ADA-145 989

DARCOM PAMPHLET

DARCOM-P 706-103

LIBRARY
RESEARCH REPORTS DIVISION
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA 93940

ENGINEERING DESIGN HANDBOOK

SELECTED TOPICS IN EXPERIMENTAL STATISTICS WITH ARMY APPLICATIONS

US ARMY MATERIEL DEVELOPMENT AND READINESS COMMAND

DECEMBER 1983

DEPARTMENT OF THE ARMY
HEADQUARTERS US ARMY MATERIEL DEVELOPMENT
AND READINESS COMMAND
5001 Eisenhower Avenue
Alexandria, VA 22333

DARCOM PAMPHLET
No. 706-103

DECEMBER 1983

ENGINEERING DESIGN HANDBOOK
SELECTED TOPICS IN EXPERIMENTAL STATISTICS WITH ARMY APPLICATIONS

TABLE OF CONTENTS

Paragraph *Page*

LIST OF ILLUSTRATIONS ix
LIST OF TABLES x
PREFACE xiii
LIST OF ABBREVIATIONS AND ACRONYMS xv

CHAPTER 1
INTRODUCTION TO CONTENTS OF THE HANDBOOK

1-1 INTRODUCTION 1-1
1-2 OVERVIEW OF THE HANDBOOK 1-1

CHAPTER 2
ERRORS OF MEASUREMENT, PRECISION, ACCURACY AND THE STATISTICAL COMPARISON OF MEASURING INSTRUMENTS

2-0 LIST OF SYMBOLS 2-1
2-1 PRELIMINARY BACKGROUND STATEMENT 2-5
2-2 INTRODUCTION AND CONCEPT FORMULATION 2-6
2-3 MEASUREMENTS WITH A SINGLE INSTRUMENT 2-8
2-4 THE SEPARATION OF PRODUCT VARIABILITY AND IMPRECISION OF MEASUREMENT WITH TWO INSTRUMENTS 2-11
2-4.1 BASIC OUTLINE AND APPROACH 2-11
2-4.2 TREATMENT OF NEGATIVE OBSERVED VARIANCES 2-15
2-4.3 A SIGNIFICANCE TEST ON IMPRECISION BASED ON TWO INSTRUMENTS 2-17
2-4.4 VARIANCES OF ESTIMATORS OF IMPRECISION OF I_1 AND I_2 2-18
2-5 THE SEPARATION OF PRODUCT VARIABILITY AND INSTRUMENT IMPRECISION WITH THREE INSTRUMENTS 2-20
2-6 SIGNIFICANCE TESTS FOR PRECISION AND ACCURACY OF TWO INSTRUMENTS 2-23
2-6.1 PRELIMINARY COMMENTS ON SIGNIFICANCE TESTS FOR TWO INSTRUMENTS 2-23
2-6.2 TEST OF WHETHER $\sigma_{e_1} = \sigma_{e_2}$ (PRECISION COMPARISON) 2-24
2-6.3 TEST OF WHETHER $\beta_1 = \beta_2$ (ACCURACY TEST) 2-25
2-6.4 LARGE SAMPLE TEST OF WHETHER σ_{e_1} OR σ_{e_2} EQUALS ZERO 2-27
2-6.5 TEST FOR WHETHER $\sigma_{e_2} = k\sigma_{e_1}$ AND SHUKLA'S TEST 2-27
2-7 SIGNIFICANCE TESTS FOR THREE INSTRUMENTS 2-29
2-7.1 INTRODUCTORY REMARKS 2-29
2-7.2 THREE-INSTRUMENT TEST OF WHETHER $\sigma_{e_1} = \sigma_{e_2}$ 2-30

TABLE OF CONTENTS (cont'd)

Paragraph		Page
2-7.3	THREE-INSTRUMENT TEST OF WHETHER $\sigma_{e_2} = k\sigma_{e_1}$ (SHUKLA'S TEST)	2-30
2-7.4	JUDGMENT PROCEDURES FOR TESTING A THIRD INSTRUMENT	2-32
2-8	CONFIDENCE BOUNDS ON THE UNKNOWN PRECISION AND ACCURACY PARAMETERS, AND ALLIED ACCOMPLISHMENTS	2-34
2-8.1	CONFIDENCE BOUNDS ON $(\beta_1 - \beta_2)$ FOR TWO INSTRUMENTS	2-34
2-8.2	CONFIDENCE BOUNDS ON $[\beta_3 - (\beta_1 + \beta_2)/2]$ FOR THREE INSTRUMENTS	2-34
2-8.3	PRELIMINARY COMMENTS ON CONFIDENCE BOUNDS FOR PRECISION PARAMETERS	2-34
2-8.4	CONFIDENCE BOUNDS ON PRECISION PARAMETERS FOR TWO INSTRUMENTS	2-35
2-8.5	CONFIDENCE BOUNDS ON PRECISION PARAMETERS FOR THREE INSTRUMENTS	2-37
2-8.5.1	Confidence Bounds on $\sigma_{e_3}^2/[(\sigma_{e_1}^2 + \sigma_{e_2}^2)/2]$	2-37
2-8.5.2	Simultaneous Confidence Bounds On All Unknown Precision Parameters	2-37
2-8.5.3	Duplicate Measurements With One of Two Instruments and Allied Results	2-38
2-8.5.4	Shukla's Three-Instrument Bounds for $\sigma_{e_2}/\sigma_{e_1}$	2-39
2-9	MEASUREMENTS WITH A GENERAL NUMBER $N \geq 3$ OF INSTRUMENTS	2-40
2-10	INTERLABORATORY TESTING FOR PRECISION AND ACCURACY STUDIES	2-43
2-11	THE HIERARCHY OF CALIBRATIONS AND THE ACCUMULATION OF ERROR	2-46
2-12	ADDITIONAL DISCUSSION OF FUNDAMENTALS OF MEASUREMENT	2-49
2-13	SUMMARY	2-50
	References	2-51
	Bibliography	2-52

CHAPTER 3
PROCEDURES FOR DETECTING OUTLYING OBSERVATIONS

3-0	LIST OF SYMBOLS	3-1
3-1	INTRODUCTION	3-3
3-2	PRELIMINARIES AND MATHEMATICAL BOUNDS OF INTEREST	3-4
3-2.1	DESIGNATION OF THE SAMPLE	3-4
3-2.2	BOUNDS FOR THE RATIO OF THE SAMPLE RANGE TO THE SAMPLE STANDARD DEVIATION	3-5
3-2.3	BOUNDS FOR THE RESIDUALS OR DEVIATIONS FROM THE SAMPLE MEAN	3-6
3-3	SOME RELATIONSHIPS AND SAMPLING DISTRIBUTIONS FOR SAMPLES OF SIZE TWO OR THREE	3-8
3-3.1	RELATION BETWEEN THE RANGE AND STANDARD DEVIATION FOR A SAMPLE OF SIZE TWO	3-8
3-3.2	THE RANGE FOR SAMPLES OF SIZE THREE AND PROPERTIES OF THE TWO CLOSEST OF THREE OBSERVATIONS	3-9
3-4	BASIS OF STATISTICAL CRITERIA FOR OUTLIERS	3-11
3-5	RECOMMENDED OUTLIER DETECTION CRITERIA FOR SINGLE SAMPLES	3-12
3-5.1	TESTS FOR EITHER THE HIGHEST OR LOWEST OBSERVATION	3-12

TABLE OF CONTENTS (cont'd)

Paragraph		*Page*
3-5.2	DIXON'S CRITERIA	3-16
3-5.3	OUTLIER TEST FOR SMALLEST AND LARGEST OBSERVATIONS	3-18
3-5.4	SIGNIFICANCE TESTS FOR THE TWO HIGHEST OR THE TWO LOWEST OBSERVATIONS	3-21
3-5.5	SIGNIFICANCE TEST FOR DETECTING SEVERAL OR MANY OUTLIERS	3-26
3-5.5.1	Preliminary Comments	3-26
3-5.5.2	The Tietjen and Moore Tests	3-26
3-5.5.3	The Rosner and Hawkins Multiple Outlier Detection Procedures	3-28
3-5.5.4	The Skewness and Kurtosis Tests for Outliers	3-37
3-6	RECOMMENDED OUTLIER TESTS USING INDEPENDENT STANDARD DEVIATION ESTIMATORS	3-40
3-7	RECOMMENDED CRITERIA FOR KNOWN STANDARD DEVIATION	3-51
3-8	THE WILK-SHAPIRO STATISTICAL TEST FOR NONNORMALITY	3-54
3-9	PROBABILITY PLOTS AND GRAPHICAL TECHNIQUES	3-54
3-10	ADDITIONAL COMMENTS AND GUIDELINES	3-55
3-11	SUMMARY	3-56
	References	3-56
	Bibliography	3-58

CHAPTER 4
SELECTED TOPICS IN ESTIMATION, THE COMMON STATISTICAL TESTS OF SIGNIFICANCE, AND THE CHOICE OF PERCENTAGE POINTS

4-0	LIST OF SYMBOLS	4-1
4-1	INTRODUCTION	4-5
4-2	PRELIMINARY REMARKS ON SAMPLING A SINGLE NORMAL POPULATION	4-6
4-2.1	THE SAMPLE MEAN AND STANDARD DEVIATION	4-6
4-2.2	THE SAMPLE MEAN DEVIATION	4-8
4-2.3	THE SAMPLE RANGE	4-10
4-2.4	BIASED OR UNBIASED ESTIMATORS AND EFFICIENCY	4-12
4-3	SOME MOMENT PROPERTIES	4-14
4-4	THE CHI-SQUARE DISTRIBUTION AND SOME OF ITS USES	4-15
4-4.1	THE CHI-SQUARE DISTRIBUTION	4-15
4-4.2	CHI-SQUARE, BINOMIAL, AND POISSON DISTRIBUTION RELATIONSHIPS	4-17
4-4.3	SIGNIFICANCE TEST FOR THE SIZE OF A NORMAL POPULATION VARIANCE	4-18
4-4.4	CONFIDENCE BOUNDS ON THE UNKNOWN POPULATION VARIANCE OR STANDARD DEVIATION	4-18
4-4.5	THE APPROXIMATE CHI-SQUARE DISTRIBUTION	4-25
4-5	THE SNEDECOR-FISHER VARIANCE RATIO OR *F* DISTRIBUTION	4-26
4-6	SIGNIFICANCE TESTS FOR THE EQUALITY OF SEVERAL POPULATION VARIANCES	4-33
4-6.1	PRELIMINARY REMARKS	4-33
4-6.2	BARTLETT'S STATISTIC	4-33
4-6.3	COCHRAN'S STATISTIC	4-34
4-6.4	HARTLEY'S STATISTIC	4-34
4-6.5	CADWELL'S STATISTIC	4-34

TABLE OF CONTENTS (cont'd)

Paragraph		*Page*
4-6.6	BARTLETT AND KENDALL'S STATISTIC	4-36
4-6.7	COMPARISONS OF THE TESTS OF HOMOSCEDASTICITY	4-38
4-6.8	FURTHER STUDIES ON HOMOSCEDASTICITY	4-38
4-7	STUDENT'S *t* DISTRIBUTION	4-40
4-7.1	INTRODUCTION	4-40
4-7.2	CONFIDENCE BOUNDS ON THE UNKNOWN NORMAL POPULATION MEAN	4-41
4-7.3	STUDENT'S *t* TEST FOR TWO NORMAL SAMPLES	4-43
4-7.3.1	Student's *t* for the Case $\sigma_1 = \sigma_2$	4-44
4-7.3.2	The Behrens-Fisher Problem ($\sigma_1 \neq \sigma_2$)	4-45
4-8	INTRODUCTORY DISCUSSION OF DESIGN AND ANALYSIS OF EXPERIMENTS	4-49
4-9	COMBINATION OF OBSERVED TAIL AREA PROBABILITES FOR INDEPENDENT EXPERIMENTS	4-56
4-10	THE CHOICE OF SIGNIFICANCE LEVELS FOR MULTIPLE TESTS	4-57
4-11	SOME FURTHER COMMENTS	4-59
4-12	SUMMARY	4-59
	References	4-60
	Bibliography	4-63

CHAPTER 5
INTRODUCTION TO SOME MODERN ANALYSES OF CONTINGENCY TABLES

5-0	LIST OF SYMBOLS	5-1
5-1	INTRODUCTION	5-3
5-2	SAMPLING A SINGLE BINOMIAL POPULATION WITH A SAMPLE OF SIZE *n*	5-4
5-3	THE 2×2 CONTINGENCY TABLE WITH EMPHASIS ON COMPARING TWO BINOMIAL POPULATIONS	5-9
5-3.1	THE GENERAL 2×2 CONTINGENCY TABLE	5-9
5-3.2	THE FISHER EXACT TEST	5-10
5-3.3	THE COMPARISON OF TWO BINOMIAL POPULATIONS	5-12
5-3.4	RECENT WORK ON COMPARING TWO BINOMIAL PERCENTAGES	5-19
5-3.5	THE DOUBLE DICHOTOMY	5-31
5-3.6	INDEPENDENCE AND INTERACTION IN 2×2 CONTINGENCY TABLES	5-33
5-4	SOME DEFINITIONS OF SYMBOLS FOR GENERAL CONTINGENCY TABLES	5-34
5-5	THE KULLBACK MINIMUM DISCRIMINATION INFORMATION STATISTICS	5-34
5-6	SOME RELATED TOPICS AND THE POWER OF 2×2 CONTINGENCY TABLES	5-37
5-7	THE GENERAL TWO-WAY CONTINGENCY TABLE (*r* Rows and *c* Columns)	5-37
5-7.1	INTRODUCTORY FORMULATION	5-37
5-7.2	THE CLASSICAL CHI-SQUARE ANALYSIS OF TWO-WAY CONTINGENCY TABLES	5-38
5-7.3	KULLBACK'S INFORMATION THEORY ANALYSIS OF TWO-WAY TABLES	5-40
5-8	COMMENTS ON THE ANALYSIS OF THREE-WAY AND HIGHER ORDER CONTINGENCY TABLES	5-43
5-9	LOGLINEAR ANALYSES OF CONTINGENCY TABLES	5-43

TABLE OF CONTENTS (cont'd)

Paragraph		*Page*
5-10	SUMMARY	5-44
	References	5-45
	Bibliography	5-46
	CHAPTER 6	
	LEAST SQUARES, REGRESSION, AND FUNCTIONAL RELATIONS	
6-0	LIST OF SYMBOLS	6-1
6-1	INTRODUCTION	6-4
6-2	LINEAR LEAST SQUARES OR REGRESSION FOR A DEPENDENT VARIABLE (MEASURED WITH ERROR) AND AN INDEPENDENT VARIABLE (WITHOUT ERROR)	6-5
6-2.1	GENERAL	6-5
6-2.2	THE LINE—ONE VARIABLE (y) SUBJECT TO ERROR	6-5
6-2.3	USE OF DEVIATIONS FROM THE MEAN	6-11
6-2.4	TRANSFORMATION OF ORIGINAL DATA FOR LINEAR LEAST SQUARES	6-12
6-2.5	EQUAL SPACING OF THE INDEPENDENT VARIABLE	6-13
6-3	LINEAR REGRESSION AND FUNCTIONAL RELATIONS—BOTH VARIABLES SUBJECT TO ERROR, BUT INDEPENDENT VARIABLE CONTROLLED	6-14
6-3.1	PRELIMINARIES TO ESTABLISH "FREE OF ERROR" IN INDEPENDENT VARIABLE	6-14
6-3.2	THE CONCEPT OF A CONTROLLED INDEPENDENT VARIABLE	6-15
6-4	LINEAR LEAST SQUARES WITH BOTH VARIABLES SUBJECT TO ERROR AND BOTH VARIABLES RANDOM	6-21
6-5	BIASES IN ESTIMATION AND BIASES IN SIGNIFICANCE TESTS DUE TO ERRORS IN THE INDEPENDENT VARIABLE	6-22
6-6	A CONSISTENT ESTIMATOR OF THE SLOPE IN A LINEAR REGRESSION MODEL WITH ERRORS IN BOTH INDEPENDENT AND DEPENDENT VARIABLES	6-23
6-7	THE PLANE: ONE VARIABLE z (THE DEPENDENT VARIABLE) SUBJECT TO ERROR	6-25
6-8	THE PARABOLA: ONE VARIABLE z (THE DEPENDENT VARIABLE) SUBJECT TO ERROR	6-29
6-9	THE REGRESSION OF A DEPENDENT VARIABLE (SUBJECT TO ERROR) ON THREE INDEPENDENT VARIABLES (FREE OF ERROR)	6-32
6-10	FITTING OF ORTHOGONAL POLYNOMIALS FOR THE CASE IN WHICH OBSERVED VALUES OF THE INDEPENDENT VARIABLE ARE AT EQUALLY SPACED INTERVALS	6-34
6-11	MULTIPLE REGRESSION OR THE GENERAL LINEAR MODEL	6-45
6-11.1	INTRODUCTION	6-45
6-11.2	THE GENERAL LINEAR REGRESSION MODEL	6-45
6-12	FUNCTIONAL RELATIONS AND NONLINEAR REGRESSION OR GENERALIZED LEAST SQUARES (WITH OR WITHOUT ERROR IN INDEPENDENT VARIABLES)	6-50
6-12.1	INTRODUCTION	6-50
6-12.2	THE GALLANT ALGORITHM (ERROR-FREE INDEPENDENT VARIABLES)	6-51
6-12.3	THE BRITT AND LUECKE ALGORITHM FOR ESTIMATING PARAMETERS IN NONLINEAR MODELS WITH ERRORS IN BOTH THE DEPENDENT AND INDEPENDENT VARIABLES	6-53

TABLE OF CONTENTS (cont'd)

Paragraph		*Page*
6-13	SUMMARY	6-55
	References	6-55
	Bibliography	6-56
	Appendix 6A	
	A Least Squares Application to Precision and Accuracy of Measurement	
6A-0	LIST OF SYMBOLS	6A-1
6A-1	PRELIMINARY REMARKS	6A-1
6A-2	ACCOUNT OF THE INTERNATIONAL OZONE ROCKET SONDE INTERCOMPARISON (IORI) STATISTICAL ANALYSIS	6A-2
6A-2.1	THE THREE-INSTRUMENT APPROACH (CONSTANT BIASES)	6A-2
6A-2.2	ESTIMATION WHEN INSTRUMENTAL BIASES CHANGE WITH ALTITUDE OR OZONE LEVEL	6A-6
6A-3	GENERAL COMMENT ON LINEAR REGRESSION WITH ERRORS IN BOTH VARIABLES	6A-12
	References	6A-12
	CHAPTER 7	
	ORDER STATISTICS AND APPLICATIONS	
7-0	LIST OF SYMBOLS	7-1
7-1	INTRODUCTION	7-3
7-2	THE DISTRIBUTION OF THE LARGEST AND SMALLEST SAMPLE VALUES, THE DISTRIBUTION OF THE RANGE, AND THE *r*th ORDER STATISTIC	7-5
7-3	THE QUASI-RANGES	7-7
7-4	EXPECTED VALUES AND MOMENTS OF SAMPLE ORDER STATISTICS	7-8
7-5	LINEAR ESTIMATION OF POPULATION PARAMETERS OR MOMENTS	7-9
7-6	DISCUSSION OF TABLES AND SOME EXAMPLES	7-9
7-7	SOME RELATIONS AND USES OF ORDER STATISTICS WITH RESPECT TO ALLIED STATISTICAL PROBLEMS	7-14
7-7.1	SOME PARTICULAR USES OF ORDER STATISTICS	7-14
7-7.2	STATISTICS OF EXTREMES	7-21
7-7.3	GUMBEL'S EXTREME VALUE DISTRIBUTION	7-23
7-7.4	ORDER STATISTICS AND OUTLYING OBSERVATIONS	7-24
7-7.5	UNIVARIATE TOLERANCE INTERVALS	7-25
7-8	ORDER STATISTICS AND THE RELATED FIELDS OF RELIABILITY AND LIFE TESTING	7-34
7-9	THE RADIAL ORDER STATISTICS AND THEIR APPLICATIONS TO TARGET ANALYSES	7-35
7-10	PARAMETER ESTIMATION FROM TRUNCATED FIRINGS AT RECTANGULAR TARGETS	7-43
7-11	PARAMETER ESTIMATION FOR TRUNCATED POISSON SAMPLES WITH MISSING ZEROS	7-49
7-12	SUMMARY	7-51
	References	7-52
	CHAPTER 8	
	DETERMINATION OF SAMPLE SIZES	
8-0	LIST OF SYMBOLS	8-1
8-1	INTRODUCTION	8-3

TABLE OF CONTENTS (cont'd)

Paragraph		*Page*
8-2	THE ROLE OF THE NORMAL DISTRIBUTION IN SAMPLE SIZE DETERMINATION	8-5
8-3	SAMPLE SIZES AND CRITERIA FOR BINOMIAL- AND POISSON-TYPE DATA	8-8
8-3.1	SAMPLING A SINGLE BINOMIAL OR POISSON POPULATION	8-8
8-3.2	SAMPLE SIZES TO COMPARE TWO BINOMIAL OR TWO POISSON POPULATIONS	8-16
8-4	SAMPLE SIZES FOR VARIANCE ESTIMATION AND COMPARISONS	8-18
8-4.1	SAMPLING A SINGLE NORMAL POPULATION TO ESTIMATE SIGMA	8-18
8-4.2	CHI-SQUARE SAMPLE SIZES FOR CONTINGENCY TABLES OR FOR CURVE FITTING	8-21
8-5	SAMPLE SIZES FOR COMPARING TWO NORMAL POPULATIONS VARIANCES	8-24
8-6	SAMPLE SIZES FOR ESTIMATION OF NORMAL POPULATION MEANS	8-28
8-6.1	SAMPLE SIZES FOR MAKING INFERENCES ABOUT THE SIZE OF A NORMAL POPULATION MEAN	8-28
8-6.2	SAMPLE SIZES FOR COMPARING THE MEANS OF TWO NORMAL POPULATIONS	8-33
8-7	POWER FUNCTION AND SAMPLE SIZES FOR THE ANALYSIS OF VARIANCE TESTS	8-36
8-8	SOME ADDITIONAL DISCUSSION ON SAMPLE SIZE DETERMINATION FOR ATTRIBUTE AND NORMAL POPULATION SAMPLINGS	8-40
8-9	SAMPLE SIZES FOR EXPONENTIAL POPULATIONS	8-41
8-10	SUMMARY	8-44
	References	8-45
	Bibliography	8-46

CHAPTER 9
SENSITIVITY ANALYSES OF QUANTAL RESPONSE TYPE DATA

9-0	LIST OF SYMBOLS	9-1
9-1	INTRODUCTION	9-3
9-2	BRIEF ANALYTICAL FORMULATION OF SENSITIVITY ANALYSES	9-4
9-3	SOME USEFUL TEST STRATEGIES	9-6
9-3.1	THE COMPLETE RUNDOWN TEST	9-7
9-3.2	THE UP AND DOWN TEST OF DIXON AND MOOD	9-7
9-3.3	THE LANGLIE ONE-SHOT STRATEGY	9-8
9-3.4	THE ROBBINS-MONRO STOCHASTIC APPROXIMATION METHOD	9-9
9-3.5	THE ONE-SHOT TRANSFORMED RESPONSE TEST STRATEGY (OSTR)	9-11
9-3.6	TRANSFORMED RESPONSE STRATEGIES FOR GENERAL n_0	9-13
9-4	ESTIMATION OF PARAMETERS	9-14
9-4.1	MAXIMUM LIKELIHOOD ESTIMATION FOR THE NORMAL MODEL	9-15
9-4.2	MAXIMUM LIKELIHOOD ESTIMATION FOR THE LOGISTIC DISTRIBUTION	9-18
9-4.3	MAXIMUM LIKELIHOOD ESTIMATION FOR THE WEIBULL MODEL	9-22
9-5	SOME REMARKS ON ALLIED WORK	9-29
9-6	SUMMARY	9-34
	References	9-34
	Bibliography	9-35

TABLE OF CONTENTS (cont'd)

Paragraph		*Page*
	Appendix 9A	
	Computer Program 9-1	9A-1
	Appendix 9B	
	Computer Program 9-2	9B-1
	CHAPTER 10	
	THE ROLE OF THE STATISTICIAN IN SCIENTIFIC MODEL BUILDING: ILLUSTRATED FOR THE LIMIT VELOCITY PROBLEM	
10-0	LIST OF SYMBOLS	10-1
10-1	INTRODUCTION	10-1
10-2	DESCRIPTION OF THE PHYSICAL AND STATISTICAL ASPECTS OF THE PROBLEM	10-2
10-2.1	BRIEF ACCOUNT OF THE PHYSICAL AND ENGINEERING DETAILS	10-2
10-2.2	THE STATISTICAL APPROACH	10-7
10-3	DISCUSSION OF THE STATE OF THE ART OF PHYSICAL AND STATISTICAL ESTIMATION OF LIMIT VELOCITY	10-11
10-4	SUMMARY	10-12
	References	10-13
	CHAPTER 11	
	INTRODUCTION TO SELECTED TOPICS IN MULTIVARIATE STATISTICAL ANALYSIS	
11-0	LIST OF SYMBOLS	11-1
11-1	INTRODUCTION	11-3
11-2	TESTS FOR EQUALITY OF POPULATION MEANS, EQUALITY OF VARIANCES, AND EQUALITY OF COVARIANCES FOR MULTIVARIATE NORMAL DISTRIBUTIONS	11-4
11-3	SELECTED TOPICS AND APPLICATIONS OF HOTELLING'S MULTIVARIATE STUDENTIZED t RATIOS AND GENERALIZED T^2 STATISTICS	11-10
11-3.1	HOTELLING'S GENERALIZATION OF THE STUDENT-FISHER t RATIOS	11-10
11-3.2	HOTELLING'S GENERALIZED T^2 STATISTICS	11-13
11-4	SUMMARY	11-36
	References	11-37
	INDEX	

LIST OF ILLUSTRATIONS

Figure No.	*Title*	*Page*
2-1	Schematic Representation of Hierarchies of Military Standards Laboratories Using National Bureau of Standards Calibration Services	2-47
4-1	95% Confidence Limits for Each Proposal	4-56
5-1	Confidence Limits for p in Binomial Sampling, Given a Sample Fraction r/n	5-6
6-1	Residual Velocity vs Striking Velocity of Projectiles	6-17
7-1	Graphs of Population Coverage for the Tolerance Level $\beta = 0.90$	7-27
7-2	Graphs of Population Coverage for the Tolerance Level $\beta = 0.95$	7-28
7-3	Graphs of Population Coverage for the Tolerance Level $\beta = 0.99$	7-29
7-4	Schematic Diagram of Target and Areas of Missing Rounds	7-45
8-1	Number of Tests vs Improvement in Reliability	8-15
8-2	Operating Characteristics of the χ^2-Test $\left[\chi^2 = \frac{(n-1)s^2}{\sigma^2}\right]$ For Testing $\sigma_1 = \sigma$ Against $\sigma_1 > \sigma$	8-22
8-3	Operating Characteristics of the χ^2-Test $\left[\chi^2 = \frac{(n-1)s^2}{\sigma^2}\right]$ For Testing $\sigma_1 = \sigma$ Against $\sigma_1 < \sigma$	8-23
8-4	Operating Characteristics of the F-Test $F = \frac{s_1^1}{s_2^2}$ For Testing $\sigma_1 = \sigma_2$ Against $\sigma_1 > \sigma_2$ $(n_1 = n_2)$	8-26
8-5	Operating Characteristics of the t-Test for $t = \left[\frac{\sqrt{n}(x-a)}{s}\right]$ For Testing $\mu = a$ Against $\mu \neq a$	8-30
8-6	Operating Characteristics of the Normal Test $z = \left[\frac{\sqrt{n}(x-a)}{\sigma}\right]$ For Testing $\mu = a$ Against $\mu \neq a$	8-32
8-7	Operating Characteristics of the F-Test $\left[F = \frac{s_1^1}{s_2^2}\right]$ For Testing $\sigma_1 = \sigma_2$ Against $\sigma_1 > \sigma_2$ $(n_1 = n_2,\ 2n_1 = 3n_2,\ n_1 = 2n_2)$	8-39
9-1	A Typical Up and Down Experiment	9-7
9-2	"One-Shot" Test to Determine Failure of Thermal Batteries	9-9
10-1	Plot of Typical Residual and Striking Velocities for a Penetrator Against Armor	10-4
10-2	Linear Relationship Between Specific Impact Energy and Scaled Armor Thickness	10-5

LIST OF TABLES

Table No.	Title	Page
2-1	Burning Times of 30 Powder Train Fuzes, s	2-8
2-2	Fuze Burning Times and Differences in Seconds	2-14
2-3	Nonnegative Variance Estimates (The Two-Instrument Case)	2-15
2-4	Estimates of Precision of Measurement on Three Simultaneous Velocity Measurements of the Fotobalk, Counter, and Terma Chronographs	2-26
2-5	Simultaneous Velocities of the Fotobalk, Counter, and Terma Chronographs With Test vs Standard Comparative Data on Each of Twelve Successive Rounds, m/s	2-33
2-6	Values of K and M Which Yield $(1 - 2\alpha)$ Confidence Regions When Used in Conjunction With Equations 2-83 Through 2-85	2-36
2-7	One-Way ANOVA Classification For Lead in Gasoline	2-44
2-8	ANOVA Table	2-45
2-9	Average Levels of the Different Laboratories	2-46
3-1	Critical Values for T (One-Sided Test of T_1 or T_n) When the Standard Deviation is Calculated From the Same Sample	3-13
3-2	Dixon Criteria for Testing of Extreme Observation (Single Sample)	3-17
3-3	Critical Values for w/s (Ratio of Range to Sample Standard Deviation)	3-19
3-4	Critical Values for $S^2_{n-1,n}/S^2$ or $S^2_{1,2}$ for Simultaneously Testing the Two Largest or Two Smallest Observations	3-22
3-5	Critical Values for E_k	3-28
3-6	Critical Values for L_k	3-31
3-7	Percentage Points of Rosner's RST Many Outlier Test Statistics R_1 and R_2	3-35
3-8	Percentage Points of Rosner's RST Many Outlier Test Statistics R_1, R_2, and R_3	3-36
3-9	Percentage Points of Rosner's RST Many Outlier Test Statistics R_1, R_2, R_3, and R_4	3-37
3-10	Percentage Points of Hawkins' E_K	3-38
3-11	Significance Levels for $\sqrt{b_1}$	3-41
3-12	Significance Levels for b_2	3-41
3-13	Critical Values for T' When Standard Deviation s_ν is Independent of Present Sample	3-42
3-14	Upper and Lower Limits for 1% Point of Distribution of $d = \max\left[\left(\frac{x_n - \bar{x}}{s}\right), \left(\frac{\bar{x} - x_1}{s}\right)\right]$	3-44
3-15	Percentage Points of Hawkins' B	3-46
3-16	Standardization of Sodium Hydroxide Solutions as Determined by Plant Laboratories	3-49
3-17	Analysis of Variance for the Data of Table 3-16	3-51
3-18	Analysis of Variance Omitting Laboratories 10 and 12	3-51
3-19	Critical Values of $T'_{1\infty}$ and $T'_{n\infty}$ When the Population Standard Deviation σ is Known	3-52
3-20	Star Plate Measurements, μm	3-53
4-1	Values of k in $\sqrt{\Sigma(x_i - \bar{x})^2}/k$ to Obtain Unbiased Estimates of Normal Population σ	4-7
4-2	Values of a_1 and a_2 Such That $a_1 s'$ and $a_2 s$ are Unbiased Estimates of the Normal Population Standard Deviation	4-9
4-3	Mean Values and Standard Deviations of the Sample Mean Deviation	4-10
4-4	Mean Values and Standard Deviations of the Sample Range w	4-11
4-5	Percentiles of the χ^2 Distribution	4-19
4-6	Divisors for the Confidence Interval About Normal Population Variance of Minimum Length	4-21

LIST OF TABLES (cont'd)

Table No.	Title	Page
4-7	Divisors for Neyman's "Shortest" Unbiased Confidence Interval for Normal Population Variance	4-23
4-8	Percentiles of the F Distribution	4-29
4-9	Impact Points	4-33
4-10	Exact Bartlett Critical Values	4-35
4-11	Hartley's Statistic; Percentage Points of the Ratio, s^2_{max}/s^2_{min}	4-37
4-12	Sample Variances	4-39
4-13	Percentiles of the t Distribution	4-42
4-14	Scott and Smith's t Approximation — (95% Level)	4-43
4-15	Observed Fuze Times	4-48
4-16	Symbolic Matrix—Grades From N Raters for K Research Proposals	4-52
4-17	Analysis of Variance of Data From Table 4-16	4-53
4-18	Scores for Six WOW Proposals by Five Raters	4-55
4-19	Analysis of Variance of Scores for WOW Proposals	4-55
4-20	Confidence Levels for Mean Scores for Six Proposals	4-55
5-1	Confidence Limits for the Expectation of a Poisson Variable	5-8
5-2	The General 2×2 Table	5-9
5-3	Exact P Values and Exact 95% Confidence Limits for Differences in Proportions in Percent (100 Δ), Ratios of Proportions R, and Odds Ratios ψ	5-16
5-4	2×2 Contingency Tables: Test for Comparing Two Proportions	5-21
5-5	Probabilities	5-31
5-6	Double Dichotomy Table	5-31
5-7	Seventy-Four Tosses of Five Coins	5-36
5-8	The General Two-Way Contingency Table	5-38
5-9	Summary of Results of IPR Questionnaire	5-39
5-10	Results of MG Firing Experiment	5-41
5-11	Two-Way Analysis of Information Table	5-42
6-1	Gravimetric Determination of Calcium in the Presence of Magnesium	6-6
6-2	Striking Velocities, Residual Velocities, and Residual Masses for 27-g Projectiles Fired Against 0.5-in. Armor Plate	6-16
6-3	Fuze Burning Times and Forward First Order Differences for Two Instruments	6-25
6-4	Ballistic Limit vs Armor Thickness and Brinell Hardness	6-28
6-5	Rifle Barrel Length vs Average Muzzle Velocity	6-31
6-6	Tables of Orthogonal Polynomials	6-37
6-7	Example of Orthogonal Polynomials	6-43
6-8	Example 6-4 Computations	6-44
6-9	ANOVA Table for Example 6-4	6-44
6A-1	Originally Measured Concentrations of Ozone in Number of Molecules per Cubic Centimeter and Instrument Differences for the Three Kreuger Instruments (UV Absorption) on Flights 249, 250, and 251	6A-7
6A-2	Analysis of Difference in Biases Between Instruments 1 and 2 Versus Altitude	6A-9
6A-3	ANOVA of Trends in Differences (Column 5, Table 6A-1) of Biases, Instrument 1 Minus Instrument 2	6A-9
7-1	Moments of Exponential, Weibull, and Gamma Populations	7-11
7-2	Expected Values of Order Statistics From $N(0,1)$	7-12
7-3	The Coefficients of the Most Efficient Linear Systematic Statistics of the Mean and Standard Deviation in Censored Samples of Sizes ≤ 10 From a Normal Population	7-15

LIST OF TABLES (cont'd)

Table No.	Title	Page
7-4	Values of $m = r + s$ Such That We May Assert With Confidence at Least β That 100_γ Percent of a Population Lies Between the rth Smallest and the sth Largest of a Random Sample of n From That Population (Continuous Distribution Function Assumed)	7-31
7-5	Confidence β With Which We May Assert That 100 γ Percent of the Population Lies Between the sth Largest and rth Smallest of a Random Sample of n From That Population (Continuous Distribution Assumed)	7-32
7-6	Tolerance Factors for Normal Distributions	7-33
7-7	Means and Standard Deviations of the Ordered Radii in a Sample of n From a Circular Normal Distribution	7-37
7-8	Variances and Covariances of the Ordered Radii in a Sample of n From a Circular Normal Distribution	7-39
7-9	Computer Program for Truncated Bivariate Normal Target Firings	7-46
7-10	Blue Tanks With One or More Hits and Observed Frequency	7-51
8-1	95% and 10% Probability Levels for the Poisson Distribution	8-11
8-2	Lower 95% Confidence Bounds on Reliability Based on Zero Failures in n Trials	8-14
8-3	Values of m, n, and θ for Best Power When $mn = 24$	8-38
9-1	OSTR Test for $n_0 = 3$, TMP = 0.7937	9-13
9-2	Characteristics of Some Transformed Response Strategies	9-14
9-3	Results of Primer Sensitivity Drop Test	9-20
9-4	Sensitivity Analysis of Projectile Initiation	9-25
9-5	Weibull Parameter Estimates	9-26
9-6	Weibull Safe-Separation Distance Percentiles (Nondetonation)	9-27
9-7	Estimated Detonation Probabilities	9-28
9-8	Optimum y and p Values for Minimum Variance Estimation of y_0	9-32
10-1	Comparison of Limit Velocity Estimation Methods	10-11
11-1	5% and 1% Points of L_{mvc} and L_{vc} for $k = 2$ and $k = 3$	11-7
11-2	5% and 1% Points of L_m	11-8
11-3	Vertical Deviation From Aim Point of First and Second Bullets Fired in Rapid-Fire Groups of Three Rounds With M16 Rifle	11-9
11-4	Upper 1% and 5% Probability or Significance Levels for Hotelling's Generalized T^2 Statistics (Bivariate Case)	11-17
11-5	Table of Coefficients for Approximating 1% and 5% Upper Probability Levels for Hotelling's Generalized T^2 Statistics	11-31
11-6	Range and Deflection Impact Positions for New and Old Artillery Projectiles	11-33

PREFACE

The continuing demand for and growth of statistical analyses in Army experimentation and applications of all kinds has resulted in a large number of special analytical techniques that are now widely used. The theory of many of the statistical techniques of special interest has been investigated systematically during the last 40 yr or so. Some of the statistical analyses of original Army interest have found their way into the broad statistical literature and, recently, into some of the university curricula. Naturally, courses in statistics taught in the universities form a strong basis for direct applications to many Army research and development efforts. As is widely recognized, the field of general statistics is indeed now an interdisciplinary science, affecting even our daily lives, and it devolves quite naturally that some special statistical procedures and experimentation guidelines would play a central role in a number of Army analytical endeavors. The need, therefore, to record and illustrate many of the well-developed statistical techniques has led to the desirability of publishing a number of engineering type handbooks on the subject of experimental statistics.

In 1962 and 1963, the US Army published five Engineering Handbooks (AMCP 706-110, -111, -112, -113, and -114) on experimental statistics, which have found extensive use and also are widely referenced in both Government and industrial activities. Our Chapter 1 gives the titles of these five volumes, along with an introductory description of the present handbook. In the intervening 20 yr or more since the publication of the AMCP 706-110 through 114 series of handbooks, much additional research in mathematical statistics has been accomplished, and some unique applications to Army problems have been found to be highly useful. Accordingly, a considerable amount of upgrading of the original material, along with some rather extensive efforts to round out and record most of the recent statistical attainments, was necessary. It is for such reasons that the present handbook has been developed.

We have endeavored to cover in considerable detail some of the topics in such fields of interest as precision and accuracy of measurement procedures, outlier detection, least squares and regression, order statistics, sample size determination and sensitivity analysis, while also including more or less supplementary coverage of techniques that have been thoroughly investigated in theory and practice or recorded in reputable current references. Topics were selected for the handbook to address the various inquires received over the past 30 yr relative to statistical problems. Hopefully, we have attained some balance in this undertaking and provided a useful compendium of some specially selected analytical procedures. It is realized that many statistical techniques not fully covered herein will no doubt find their way into future Army practice; a specific cutoff date for a handbook dictates the particular selection of topics that can be included. Nevertheless, the techniques we have included should be of general use for many years to come. In fact, it is visualized that some of our selected subjects will come into prominence not only in Army applications but also in industrial, engineering, and research pursuits as well. In any event, it is hoped that we have provided a sound basis for future applications and have indicated some areas for further research. It is believed that the reader will find many references in this volume which should prove of value in his Army statistical endeavors.

The development of this book is almost wholly the work of Dr. Frank E. Grubbs, formerly Chief Operations Research Analyst of the US Army Ballistic Research Laboratories. Dr. Grubbs was in fact engaged in much of the Army's statistical programs during the years 1941 to 1981. Indeed much of his research in mathematical statistics, which has been found extensively applicable in Army and industrial problems, is recorded in this handbook. We are much indebted to the US Army Materiel Systems Analysis Activity (AMSAA) and the US Army Ballistic Research Laboratory (BRL) for providing support during the preparation of this handbook.

The US Army DARCOM policy is to release these Engineering Design Handbooks in accordance with DOD Directive 7230.7, 18 September 1973. Procedures for acquiring Handbooks follow:

a. All Department of Army (DA) activities that have a need for Handbooks should submit their request on an official requisition form (DA Form 17, 17 January 1970) directly to:

Commander
Letterkenny Army Depot
ATTN: SDSLE-SAAD
Chambersburg, PA 17201.

"Need to know" justification must accompany requests for classified Handbooks. DA activities will not requisition Handbooks for further free distribution.

b. DOD, Navy, Air Force, Marine Corps, nonmilitary Government agencies, contractors, private industry, individuals, and others—who are registered with the Defense Technical Information Center (DTIC) and have a National Technical Information Service (NTIS) deposit account—may obtain Handbooks from:

Defense Technical Information Center
Cameron Station
Alexandria, VA 22314.

c. Requestors, not part of DA nor registered with the DTIC, may purchase unclassified Handbooks from:

National Technical Information Service
Department of Commerce
Springfield, VA 22161.

Comments and suggestions on this Handbook are welcome and should be addressed to:

Commander
US Army Materiel Development and Readiness Command
Alexandria, VA 22333.

(DA Form 2028, Recommended Changes to Publications, which is available through normal publication channels, may be used for comments/suggestions.)

LIST OF ABBREVIATIONS AND ACRONYMS

AMSAA = US Army Materiel Systems Analysis Activity
ANOVA = analysis of variance
AP = armor piercing
ASTM = American Society for Testing and Materials
ATGM = antitank guided missile
BHN = Brinell hardness number
BL = ballistic limit
BL = barrel length
BLUE = best linear unbiased estimator
BRL = US Army Ballistic Research Laboratory
cdf = cumulative distribution function
CEP = circular error probable
C of I = center of impact
CSM = convexity, symmetry, and the maximum condition
CT = Cochran's test
D = down
DA = Department of the Army
df = degree of freedom
DOD = Department of Defense
ESD = extreme studentized deviate
F = Snedecor-Fisher F statistic
FAA = Federal Aviation Administration
GPO = Government Printing Office
HE = high explosive
IORI = International Ozone Rocket Sonde Intercomparison (Study)
IPR = in-process review
LCL = lower confidence limit
LHS = left-hand side
logit = $\ln (p/q)$
MCS = minimum chi-square
MD = mean deviation
MDIS = minimum discrimination information statistic
MG = machine gun
ML = maximum likelihood
MMBF = mean-miles-between-failures
MS = mean square
MSE = mean square error
MV = muzzle velocity
NASA = National Aeronautics and Space Administration
OC = operating characteristic (curve)
OSTR = one-shot test response
OSTR = one-shot transformed response
PAP = potassium acid phthalate
pdf = probability density function
ppm = parts per million
PTS = preliminary test of significance
R&D = research and development
RHS = right-hand side
RST = R statistics (outliers)
SS = sum(s) of squares
TI = Tchebycheff inequality
TMP = transformed median percentage
TMR = transformed median response
U = up
UCL = upper confidence limit
UMPU = uniformly most powerful unbiased

CHAPTER 1
INTRODUCTION TO CONTENTS OF THE HANDBOOK

A brief but somewhat comprehensive and explanatory view of the topics and general subject matter of the handbook is highlighted in this chapter.

1-1 INTRODUCTION

During the 1960's a series of Engineering Design Handbooks on the general subject of experimental statistics was published by the US Army. These Engineering Design Handbooks have the following pamphlet numbers and titles:

AMCP 706-	Title
110	*Experimental Statistics, Section 1, Basic Concepts and Analysis of Measurement Data*
111	*Experimental Statistics, Section 2, Analysis of Enumerative and Classificatory Data*
112	*Experimental Statistics, Section 3, Planning and Analysis of Comparative Experiments*
113	*Experimental Statistics, Section 4, Special Topics*
114	*Experimental Statistics, Section 5, Tables.*

This valuable set of handbooks on experimental statistics and related subjects has served the Army analysts quite well as an authoritative reference of useful methodology and examples. In the intervening years, however, the field of experimental statistics has moved forward at a very rapid pace, and in fact, many new and useful techniques in experimental statistics have become available. Our primary objectives in the preparation of this handbook, therefore, have been to select some of the more useful statistical techniques we believed Army analysts would require and to assemble them in a single, comprehensive volume. As would no doubt be expected, we were not able to devote the space to cover the multitude of many other desirable statistical methods—for example, extensive multivariate distribution theory (or even bivariate or trivariate weapon delivery error distributions), the estimation of (residual) dispersion from mean square successive or higher order differences, or nonparametric statistics to the extent desired. Moreover, it seemed too early to cover the use and applications of "robust" statistical estimation methods, even though some special interest has been evident in this area. Nevertheless, we consider that the topics we have covered in this handbook will represent a valuable addition to the Experimental Statistics series of handbooks—AMCP 706-110 through -114—and will either provide the analyst with useful reference material or perhaps help him with the current methodology of some of the more up-to-date advances.

1-2 OVERVIEW OF THE HANDBOOK

We have presented the topics in this handbook in a certain order to draw proper attention to application areas that are now considered mandatory for the successful, practicing experimental statistician. Thus we have not approached the general subject of Army experimental statistics in what some might regard as a logical order of elementary statistical concepts in a college- or university-type curriculum. In fact, we have long observed that the more usual college statistical courses do not even approach the need to handle or deal effectively with the formidable problems in practice—another reason for preparing this handbook. As a case in point, consider the problem of errors in measurement, precision, and accuracy of measurement. It is certainly of considerable interest to know in much detail just how well errors of measurement are controlled; otherwise the observations taken in an experiment could lead to entirely wrong conclusions and inferences. Hence perhaps the prime objective in experimental work is the assurance that the measurements taken will be of proper quality. It is for this reason that we devote attention first in Chapter 2 to the statistical treatment of errors of measurement, precision, and accuracy problems. We attempt to define, provide methods of estimation, and illustrate by actual example these very elusive concepts in Chapter 2. Moreover, coverage in Chapter 2 includes the known, key statistical tests of significance, which are useful in comparing population parameters of the precision and accuracy measures. In

dealing with these problems of precision and accuracy of measurement, it is necessary to discuss the hierarchy of calibration echelons to the top, or the National Bureau of Standards, and the probable accumulation of error through such channels. Finally, the use of interlaboratory studies of measurement procedures and test methods, or "round-robin" tests must be considered. Thus we have given an introduction to these practices and procedures in Chapter 2 also. With suitable knowledge of the precision and accuracy of our measurement procedures, we are ready to discuss the next logical topic in statistical practice, namely, the analysis and treatment of outliers.

Chapter 3 gives an account of the statistical tests that are rather widely used in current applications to identify and to isolate outlying observations in samples. The so-called "outliers" that often appear in experimental work could be due to errors of measurement, recording errors, or just plain mistakes, but they also could reflect the true characteristics of the population one is actually sampling. Thus the basic problem is to develop the more useful statistical tests that will lead almost unerringly to the separation of true outliers from the actual characteristics of the population sampled, i.e., the physical environment. For a systematic and comprehensive treatment of the outlier detection problem in Chapter 3, we give the more efficient statistical procedures for isolating either a single high or single low anomalous observation, or either the two highest or the two lowest sample values, and also some rules for judgment of the lowest and the highest observations simultaneously. For small samples these particular cases are met very frequently in many practical situations. We then proceed to discuss in some detail the detection of many outliers (more than two) or, that is, the likelihood of much unacceptable heterogeneity in the sample of observations. Several multiple outlier detection procedures are given, and pertinent practical examples are illustrated. Since our interest lies in the realm of making sound conclusions and inferences based on the statistical analysis, the methods of Chapters 2 and 3 become of fundamental importance in helping to assign the likely causes of questionable variations.

Hence Chapters 2 and 3 have been placed first to call close attention to and also to provide the Army statistical analyst with a solid background for handling and assessing errors of measurement and the possible effect of outliers in important practical applications. We believe that this approach to modern day statistical analyses leads us with much assurance to the proper handling of the many special or selected techniques discussed herein, which currently are required in many applied Army investigations.

There is a variety of special statistical topics, that have come to light over the years, and, as a matter of fact, have been found to be of much particular interest to the practicing statistician. Moreover, it seemed very highly desirable to bring these topics together in a single chapter, which we have done in Chapter 4. Such topics include, for example, some elementary account of basic estimation techniques—particularly approximate unbiased estimation of the population standard deviation for samples from a normal population, the concepts of efficiency and mean square error, some updating of the common statistical tests of significance, and some points on the choice of significance levels for multiple tests. In recent years there have been some advances in the development of approximate statistical procedures for some of the significance tests, and for many or most practical applications such techniques may just as well be used. In the Student type t tests for comparing normal population means, the use of $(n - 3)$ instead of $(n - 1)$ degrees of freedom (df) as a divisor of the sum of squares leads to a t statistic that is very nearly normally distributed. Hence the table of standardized normal deviates—instead of the usual t table—may be used in practice, and in fact, only a normal percentage point must be remembered! Moreover, this development extends rather well to both the two-sample t test and the Behrens-Fisher problem for comparing two normal population means for which the variances are not equal. Clearly, such suitable, approximate techniques could well promote wider practical applications because the rigorous handling of only the exact tests has been intractable. Along with the common statistical tests of significance *et al.*, there seemed to be some value in recording the principles of establishing confidence bounds on the unknown normal population sigma or standard deviation, including a discussion of Neyman's shortest unbiased confidence bounds. These topics are covered in Chapter 4.

Since the applied statistician often must compare the relative size of more than two normal population sigmas, up-to-date coverage of significance tests for the equality of several population variances must be approached. Hence homoscedasticity tests, such as that of Bartlett, Cochran, Hartley, Cadwell, and Bartlett and Kendall, are highlighted in Chapter 4.

The design and analysis of planned experiments using statistical experience now extend over such a wide area that we cannot go into such developments and accomplishments in this handbook. Also many excellent textbooks on the general subject are now widely available. Nevertheless, we considered it desirable to discuss a rather frequently appearing problem of comparing subjective type judgments in much Army work. Our analysis of variance technique used here concerns the rating and ranking of research and development proposals by a panel of "experts"; many similar applications could be made elsewhere. As the final subject of Chapter 4, we discuss the choice of significance levels for multiple type tests. There are often cases that involve a series of significance tests, and in the end one desires to guarantee a given or prestated level of significance.

As would be expected, many Army statistical applications involve the comparison of two unknown binomial population parameters or some analyses of count or cross-classified categorical data. One of the most frequent and classical problems concerns the analysis of 2×2 comparative trials, or two-way contingency tables, especially the 2×2 table of count data. In Chapter 5 we have tried to give some of the more relevant background concerning the analysis of 2×2 contingency tables by using the classical normal approximations and the chi-square analysis equivalent test. As has been recognized since the 1940's, one has to consider both the possibilities of fixed and variable marginal totals with the classical comparison of two binomial population parameters imbedded in such treatments. We follow the basic work of Barnard and Pearson in this endeavor and attempt to give much assurance to the fact that the normal approximation is normally quite satisfactory. Since there has been much confusion in the past concerning both the interpretation and the statistical analysis of contingency tables, we have tried to develop and present the material in an order and fashion the Army analyst can follow and remember. This means that for the frequently used 2×2 table the comparison of two binomial population parameters or proportions appears to be of some central importance. This case, therefore, is treated rather extensively, and some Army type applications are given.

During the past 20 yr or so, there have been some developments toward "different" approaches to the analysis of contingency tables, including the information theory approach and the loglinear model. Consequently, we have included some discussion of both of these approaches, even though somewhat limited in scope, while adhering to the belief that analyses should treat the original, observed count data without any transformation of scale. We must note, however, that the use of the loglinear model leads to linearization of the data and hence likens this approach to the well-known analysis of variance (ANOVA) of statistically designed experiments, such as two-way classifications or layouts of randomized blocks.

Due to the demand for statistical analyses arising from diverse applications, readers should be aware that least squares, regression, and the fitting of functional relations represent some of the most important topics to be covered in any handbook of this kind. Moreover, practical applications now require more than just a "routine fit" as is sometimes presented in statistical textbooks. In fact, in line with the principles of Chapter 2, present-day analysts should have profound appreciation for the existence and size of errors of measurement and whether or not the dependent variable is sufficiently "free of error" or otherwise deserves some special treatment. Consequently, Chapter 6 has been written with such problems in mind for attacking least squares. Also for these reasons the very first problem or example illustrated is approached from the standpoint of whether the assumptions and the fitted linear model are valid. In this way one can perform least squares in such a manner as to have great assurance and confidence for his analytical judgments.

Although statisticians, using the fitted equation statistics, have long determined confidence intervals for specific values, an important result of Henry Scheffé that covers multiple confidence statements about and for the whole least squares line has too long been overlooked. Therefore, Scheffé's theory for the regression line and its practical benefits are stressed. Also the important result of Berkson, which points out that when the experimenter presets and aims for "controlled" values of the independent variable, the ordinary least squares line involving y on x may be fitted in the normal manner as for x free of error. We go to some effort in Chapter 6, therefore, to select and exhibit those regression topics that may be of most importance in practice.

Although physical scientists have always faced the least squares case involving "errors" in both variables, i.e., the dependent and the independent variables, it is only in recent years that the statistician has

developed an appropriate treatment of this problem. Hence the "errors in both variables" case is discussed very thoroughly, and modern approaches for use are presented. Also we stress in Chapter 6 the comparison between the fitting of an appropriate physical model on one hand and that of a polynomial on the other. The value of the physical model is demonstrated by using a problem in penetration mechanics.

The fitting of a dependent variable on several independent variables is presented in a rather simple computational manner. The use of orthogonal polynomials for equally spaced values of the abscissa is stressed in connection with the analysis of variance (ANOVA) table, which uses a Snedecor-Fisher F test for a stopping rule. A very unique example, applying Chapter 2 principles, is also given.

The need for analyses of the ordered observations in a sample, as contrasted to observations in the order taken, has deserved much special attention in recent years. This is due to the fields of life testing and reliability, where the lifetimes of articles occur naturally in increasing order and such tests may be stopped before all articles fail; or the existence of outliers in samples; or some rounds fired at a target that miss it, etc.; and for which unbiased estimation of population parameters is required. Indeed, the rather incontrovertible results arising from estimation through the use of sample order statistics make their applications very attractive for their efficiency is surprisingly high. Thus Chapter 7 attempts to present an introductory account of some of the principles involved in the analysis of sample order statistics for purposes of inference. Our interest in order statistics concentrates on distributions of largest and smallest values in the sample, the sample range or largest minus smallest values, the quasi-ranges, expected values of the sample order statistics and their moments, efficient linear estimation of population parameters, the statistics of extremes and Gumbel's extreme value distribution, some relationships between order statistics and outliers, the radial order statistics as applied to target analyses, the analysis of truncated samples from firings at rectangular targets, and parameter estimation for truncated Poisson samples with missing zero occurrences. The last-named application applies, for example, to the analysis of combat records about tank engagements for which the number of misses is naturally never known but the number of tanks having one hit, two hits, or more is identifiable.

In terms of order statistics, several distributions come into importance in applications. These include the normal, the exponential, and the Weibull distributions. In Chapter 7 we illustrate the use of order statistic theory by a number of examples that illustrate the versatility of this analytical tool.

Perhaps the most ubiquitous requirement of a statistical character among physical scientists and others concerns that of selecting the right sample size. In fact, the almost universal question is invariably, "What sample size do I need?". This question is certainly a very simple one but often like others requires some qualification, to say the least! The determination of sample size is not only or strictly a statistical problem, but it may be a physical or engineering one as well or even an economical one since as so often one "gets only what he pays for". In some cases the sample size is limited by just what is actually available for test, in which case the design of the test might well come into play. On the other hand, the statistical determination of sample size represents an important activity because there must be some control of the risks of erroneous judgments. That is to say, for example, that we would like to keep the "Type I" error of rejecting a "good product" and the "Type II" error of accepting a "bad product" both down to a minimum. Perhaps it is easy to see then that the determination of sample size is very dependent on the variability of the population to be sampled, or, that is, the population standard deviation. If this sigma is small, the sample size will ordinarily be smaller than if the sigma were large. Also the choice of sample size will depend very much on just how close we desire to be near the population parameter—i.e., mean, standard deviation, etc. Clearly, if we desire that the sample mean be the same as the population mean, the sample size and the population size must be equal, or very nearly so. What we are also saying in effect is that sample size determination will depend on the particular difference we would like to be able to detect and the width of the confidence interval within which we would like the population parameter to lie. Hence there are a number of ways of framing questions concerning sample size determination, and the approach must be selected with some care. Moreover, once the appropriate approach has been selected, the sample size must not be so large as to be impracticable—a final requirement.

It might be said that we more or less focus on two approaches having some practical value for the determination of sample size. The first of these revolves around either establishing a difference of practical importance or a deviation from the population parameter we would like to detect and then finding the

sample size for the significance test that will show statistical significance for the probability level also selected. This particular approach is often used because it is not difficult for the practicing engineer or physical scientist to formulate and to apply. The second, and perhaps more difficult, approach for the practitioner is to formulate the problem in terms of just what is acceptable or desirable and what level of quality, etc., is not, then to determine the risks one might be willing to take in these two judgments, and finally to obtain the sample size that guarantees these attainments. In this way we are controlling the risks of erroneous judgments. In Chapter 8 we discuss both approaches in an appropriately detailed manner for the more common statistical tests of significance, and we illustrate the principles by a number of practical examples.

The determination of sample size(s) is recorded for sampling a single binomial population or comparing two binomial populations (or Poisson distributions); the testing for high reliability; the estimation and comparison of normal population variances; the estimation and comparison of normal population means, and the normal populations; contingency tables and curve fitting; and a brief account of sample sizes for analysis of variance type problems. Every effort is made to keep the sample size equations as simple as possible, and particular attention is given to the use of the normal approximations by showing their accuracy. Thus the practicing statistician should find much use for Chapter 8.

Long before statistical techniques were applied in depth to industrial- and engineering-type problems, there existed a need to use probabilistic methodology in bioassay problems or "dosage response" analyses. This perhaps was especially the case since the data were of a "quantal response" type nature or an "all or nothing" response. Thus the analyst appeared to be face-to-face with an application involving a continuous scale, or "variables", treatment, but the response data were simply of an "attribute" nature, or "yes" or "no" character. For the Army the pressing need for quantal response analyses came to the forefront in connection with analyses of armor penetration studies and the mammoth effort directed toward acceptance testing of armor plate from many producers during World War II. The analytical problem is clearly seen for defeat of armor studies since, in firing projectiles at armor of a given thickness, there exists some "lower" striking velocity for which no penetrations of the plate occur, but as the striking velocity is increased, there are 10%, 20%, . . ., 50%, . . ., 90%, . . ., and finally perhaps even 100% penetrations at some "higher" velocity. Hence basically one must estimate a cumulative distribution curve, which is most often unknown, for the case where the firing of a single round results in either a nonperforation or a perforation. Moreover, it is starkly clear that firings near the levels of 0% or 100% perforations give little or no useful information! Therefore, one must also adopt an efficient strategy for conducting armor penetration tests if he is to obtain the characteristics of the "zone of mixed results". For industrial and engineering applications, this particular type of statistical problem was most often branded as a "sensitivity analysis" as contrasted to the specific bioassay procedure. Chapter 9 discusses some of the more up-to-date methods for sensitivity analyses of quantal response type data.

Since the problem in experimental testing for sensitivity analyses is that of locating the zone of mixed results and exploring it in a fashion to estimate parameters of the assumed or guessed-at distribution, the strategy of conducting the test and the related statistical analysis must go hand-in-hand. Hence, if one has to determine a low percentage point, say 1%, or a high percentage point, say 99%, then the strategy of testing should be so aimed. On the other hand, if one is primarily interested in the median, or 50%, probability level and some idea concerning the width or standard deviation of the zone of mixed results, he should avoid the end points and simply assume that the distribution is normal. For the zone of mixed results, the distributions covered in Chapter 10 include the normal, the logistic, and the Weibull models. The discussion, therefore, involves a variety of distributional shapes. Testing strategies include the complete rundown test, the "up and down" strategy of Dixon and Mood, the Langlie one-shot test strategy, the Robbins-Monro stochastic approximation method, the one-shot transformed response test strategy (OSTR), and more general transformed response strategies for extreme percentage points of the assumed distributions. The primary technique for the estimation of population parameters is Fisher's method of maximum likelihood, and some discussion of the iterative procedures is given as required. Also a number of very informative examples and computational aids add to the usefulness of Chapter 9 for Army applications.

Chapter 10 has been selected and prepared with a special purpose in mind. Our objective is to outline a rather difficult problem that can be used to indicate the contrast between the statistical approach to model development as compared to that of the physical approach and just how they might support each other. In fact, the statistician would often fare better by trying to fit the available physical models to the data before attempting to improve their applicability statistically. As it turns out, the applied or consulting statistician will be called upon to use his expertise in any number of diverse areas of emphasis, and it is unlikely that he will have immediately at hand the detailed knowledge required in each and every field or problem. Likewise, as so often happens, the physical scientist will not be sufficiently trained in statistical methodology; therefore, the best approach must be teamwork involving both viewpoints. Communication barriers have been disappearing in recent years, and proper coordination should no longer be a stumbling block since the multidisciplinary approach represents a common practice in science, technology, and engineering. We believe that such practices will be a continuing necessity.

For purposes of a convincing illustration, we have chosen the so-called "limit velocity" or "critical velocity" problem in penetration mechanics studies. The limit velocity of a target armor plate may be defined as the greatest striking velocity for which the chance of penetration is zero in statistical terms, or in physical terms it is the striking velocity for which the residual velocity is zero. Even though the reader may be aware of some similarity between Chapter 10 and the statistical sensitivity analyses of Chapter 9, there is a sharp and important difference that must be recognized. In fact, Chapter 9 is concerned with only the statistical approach or analysis of quantile response data, whereas Chapter 10 involves measurements on both a continuous and attributive scale along with the problem of determining a physical law that will give the limit velocity in terms of the armor thickness and hardness, the projectile diameter, the projectile mass, the striking velocity, the angle of striking obliquity, and other physical parameters. In other words we take up the problem of describing the role of the statistician as a team member in the activity of scientific model building or development. The requirement for coordinating the roles of the statistician and the physical scientist is discussed and amplified.

The final chapter, Chapter 11, focuses on an introduction to some selected topics in multivariate statistical analysis and theory since a number of key problems arise in connection with many Army applications of statistical methodology. For example, some weapons have circular patterns of shots, i.e., equal sigmas in the different directions, and it becomes desirable to test for "circularity". Statistical problems of this nature may be handled by using Wilks' likelihood ratio tests for determining the equality of variances, the equality of covariances, and the equality of mean values also. Usually, one is dealing with a single bivariate or multivariate sample for the problems of this type, and we give an illustration for the M16 rifle in rapid fire to indicate the nature of the application.

Chapter 11 also includes bivariate and multivariate statistical theory for comparing the results of two samples with each item of the sample having multiple characteristics. Here one often needs to compare the true covariance matrices of two bivariate or multivariate normal populations and uses the Hotelling generalized T^2 statistic, or he needs to compare the corresponding true characteristic means of two hypothesized multivariate normal populations, in which case the application of Hotelling's multivariate Studentized t statistic is required. Finally, a Hotelling generalized T^2 statistic can be used to test whether two multivariate normal samples can be considered to originate from a single multivariate normal population. These Hotelling T^2 statistics are thoroughly illustrated with an example that compares a newly designed and a standard artillery projectile.

Since many users of this handbook may have applications that will require the simultaneous use of statistical methods from several of the chapters, we have selected a comprehensive and rather extensive problem related to a study of the precision and accuracy of instrumentation for determining the stratospheric ozone concentration in the atmosphere. This statistical analysis requires the application of the principles of Chapter 2, which requires redundancy of instrumentation to estimate the imprecision of measurement of each measuring device, and along with it the application of orthogonal least squares procedures covered in Chapter 6 to model the trends in instrumental bias differences. As a result, one can develop precision and accuracy statements for the capabilities of the instruments and hence settle any error of measurement questions. This study is presented in the Appendix of Chapter 6.

CHAPTER 2

ERRORS OF MEASUREMENT, PRECISION, ACCURACY AND THE STATISTICAL COMPARISON OF MEASURING INSTRUMENTS

Precision and accuracy of measurement represent widely misunderstood terms or concepts with the result that many controversies arise in science, technology, and industrial practice. We therefore attempt to define and quantify errors of measurement, precision, and accuracy in accordance with the principles of statistics that apply so aptly to these concepts. By means of a systematic approach to the problem, precision and accuracy (or imprecision and inaccuracy) are described in an analytical manner, and the statistical techniques of estimating these parameters are given. It is found that at least two measuring instruments, taking common or the same measurements, are required to provide the needed estimates and to obtain some idea concerning the reliability of the estimates. Moreover, these principles are extended to any number of measuring instruments or laboratories engaged in measurement operations.

Many pertinent statistical tests of significance concerning the precision and accuracy (large sample or population) parameters are presented for the analyst, and procedures for establishing confidence bounds on the unknown parameters of measurement are also covered in considerable detail. These results are discussed especially for either two or three instruments, and indications of usage are given for any general number of measuring instruments.

The practice of interlaboratory testing is covered in some analytical detail, and techniques for estimating the components of variance (or the repeatability and reproducibility sigmas) are illustrated numerically.

Finally, we give an account of the hierarchy of calibration echelons or channels and present an analysis of the accumulation of error in such procedures. Many practical examples are given to illustrate the theory.

2-0 LIST OF SYMBOLS

$A = r_{uv}^2 - P$

$A_{rr} = n \sum_{i=1}^{n} r_i^2 - (\Sigma r_i)^2$ = convenient notation for n times the sum of squares about the sample mean. (Applies also to any other letter subscripts.)

a = optimum value determined by minimizing total costs of calibration laboratory hierarchy

a_0 = constant or exponent (see Eq. 2-137)

a_1 = constant or exponent (see Eq. 2-138)

$B = 2[(r_{uv}^2 - P) + (1 - P)S_{uv}/S_v^2]$

b_0 = constant or coefficient (see Eq. 2-138)

b_1 = constant or coefficient (see Eq. 2-138)

$C = r_{uv}^2 - P + (1 - P)\,[(S_u^2/S_v^2) + 2S_{uv}/S_v^2]$

$c = \sigma_{i+1}/\sigma_i = \sigma_x/\sigma_i$ = constant precision ratio at each and every calibration echelon i

D_L = lower confidence limit (see Eq. 2-90)

D_U = upper confidence limit (see Eq. 2-91)

E = error committed at a laboratory

$E(\quad)$ = expected value or large sample average of ( ), the quantity within parentheses

e = random error of measurement whose mean or expected value is zero

$\bar{e} = \sum_{i=1}^{n} e_i/n$ = sample average of the random e_i for n items

e' = total error of measurement or instrumental error, including bias and random error

$\bar{e}' = \sum_{i=1}^{n} e_i'/n$ = sample average error of measurement for n items

e_i = random error of measurement for ith item

e_{ij} = random error of measurement for the ith reading of the jth instrument, where $j = 1, 2$, etc. e_{i1} is the ith random error of measurement for I_1 concerning the ith item. The e_{ij} are assumed to be normally distributed with the zero mean and variance σ^2_{ej}.

$F = S_s^2 - S_{rs}$ for use in Shukla's technique (see Eq. 2-86)

F_0 = observed value of F

$F(n-1, n-1)$ = refers to Snedecor-Fisher F distribution with $(n-1)$ and $(n-1)$ degrees of freedom

$G = S_r^2 - S_{rs}$ for use in Shukla's technique (see Eq. 2-87)

$g_i = s_i + k^2 r_i$ for Shukla's technique

$H = t_\alpha^2 (S_r^2 S_s^2 - S_{rs}^2)/(n-2)$ (see Eq. 2-88)

H_0 = null hypothesis to be tested

H_a = alternative hypothesis

$h_i = u_i + (\delta + 1)v_i$

I_j = jth measuring instrument: $j = 1, 2, \ldots$

K = constant or factor for Thompson's confidence bounds in Eqs. 2-83 through 2-85 and Table 2-6

$K = [(S_r^2 - S_s^2)^2 - 4(S_r^2 - S_{rs})(S_s^2 - S_{rs})]^{1/2}$ = convenient parameter in Eq. 2-32

k = constant or multiplier

k = number of participating laboratories in an interlaboratory test

k = ratio of imprecisions $\sigma_{e_2}/\sigma_{e_1}$, e.g., in Eq. 2-68

ℓ = factor or constant for a lower confidence bound of Hanumara and Thompson (see Eqs. 2-95 and 2-96)

M = constant or factor for Thompson's confidence bounds in Eqs. 2-83 through 2-85 and Table 2-6

m = number of calibration echelons

m_i = number of laboratories at echelon i

N = total number of instruments, observations, runs

$N(0,1)$ = denotes a random variable that is normally distributed with zero mean and unit standard deviation or variance

n = number of measurements or sample size

n_j = number of observations in jth column

$P = t_{1-\alpha}^2/(t_{1-\alpha}^2 + n - 2)$

$p_i = r_i + s_i = \beta_1 + \beta_2 + 2x_i + e_{i1} + e_{i2}$ = sum of readings of instruments I_1 and I_2 for ith item

Q_j = particular variance of residuals, defined in Eq. 2-141, which is equivalent to the variance of errors of measurement of the jth instrument

$q = u_{i_n} + (\delta + 1)v_i$

$R = \sum_{j=1}$ = number of "runs" made with all instruments

RHS = right-hand side of

$r = \mu + e' = \mu + \beta + e$

r = observed value of a measurement for the first instrument I_1

$r_i = \sigma_x/\sigma_{e_1}$ = precision ratio

$r_i = x_i + \beta + e_i$ = observed value (measurement) for the ith reading or measurement with instrument I_1

r_i = ith measurement or reading of I_1

r_{ij} = ith reading of the jth instrument

r_{ij} = used to denote the element or cell value in the ith row and jth column of a two-way classification in the analysis of variance table (see Eq. 2-140)

$r_{ik} = \alpha_k + \beta_k x_i + e_{ik}$ = observed value or readings on ith item for "run" k

r_{ik} = ith reading of the kth instrument

r_j = number of "runs" made with instrument I_j

$r_T = \sqrt{c^2 - 1}$ = total precision ratio

$r_{xe} = S_{xe}/(S_x S_e)$ = sample correlation coefficient of the true values x and the errors of measurement e (Applies also to any other different letter subscripts, e.g., r_{xy}, r_{uv}, etc.)

$\bar{r}_{i.}$ = average of a row, i.e., averaged over the columns

$\bar{r}_{..}$ = grand average of the two-way analysis of variance table

$\bar{r}_j$ = sample mean of the readings of instrument j

$\bar{r}_{.j}$ = average of a column, i.e., averaged over the rows

$\bar{r}_{.k}$ = sample mean of the readings of instrument k

$S_e^2 = [1/(n-1)] \sum_{i=1}^{n} (e_i - \bar{e})^2$ = sample variance of the errors of measurement

$S_{e_1 e_2}$ = sample covariance of errors of measurement of I_1 and I_2

$S^2_{e_j - e_k}$ = sample variance of the differences in readings of instruments I_j and I_k

S_j = special symbol (see Eq. 2-139) used to denote the residual variance when row and column effects have been eliminated

$S_{jj} = S_j^2$ = sample variance of the readings of instrument I_j

S_{jk} = generally a sample covariance term for instrument readings of I_j and I_k (see Eq. 2-94)

$S_r^2 = S^2_{x+e_1} = \sum_{i=1}^{n} (r_i - \bar{r})^2/(n-1) = A_{rr}/[n(n-1)]$ = sample variance for instrument I_1 based on $(n-1)$ degrees of freedom. (Applies also to any other letter subscripts, e.g., S_s^2, S_u^2, S_e^2, etc.)

$S_{rs} = S_{x+e_1,\ x+e_2}$ = covariance of the readings of the first and second instruments I_1 and I_2

$S^2_{r-s} = S^2(r-s) = S_u^2 = S^2_{e_1 - e_2}$ = sample variance of difference in readings of instruments I_1 and I_2

S^2_{r+s} = sample variance of the sum of readings of instruments I_1 and I_2

S^2_{r+s+t} = sample variance of the sum of the three instrument readings for each item measured

$S^2_{\bar{r}+\bar{s}+\bar{t}}$ = sample variance of the average of the three instrument readings for each item measured

S_s^2 = sample variance of instrument I_2 based on $(n-1)$ degrees of freedom

S_{st} = covariance of the readings of instruments I_2 and I_3

S_u^2 = sample variance of the difference in readings of instruments I_1 and I_2

S_v^2 = sample variance of the difference in readings of instruments I_2 and I_3

S_w^2 = sample variance of the difference in readings of instruments I_3 and I_1

$S_x^2 = [1/(n-1)] \sum_{i=1}^{n} (x_i - \bar{x})^2$ = sample variance of the true unknown values of the characteristic or item measured

$S^2_{x+e_j}$ = sample variance of readings of the jth instrument I_j

$S^2_{x+e_1} = S_r^2$ = sample variance of the readings of the 1st instrument, for example

$S_{xe} = \sum_{i=1}^{n} (x_i - \bar{x})(e_i - \bar{e})/(n-1) = A_{xe}/[n(n-1)]$ = sample covariance of the true values x and the errors of measurement e. (Applies also to any other letter subscripts, e.g., S_{uv}, S_{xy}, etc.)

S_{xe_1} = covariance of true values and errors of measurement of I_1

S_{xe_2} = covariance of true values and errors of measurement of I_2

$S_{x+e_j,\ x+e_k}$ = sample covariance of the sum of readings of instruments I_j and I_k

$S_{x+e_j,\ x+e_k} = S_{x+e_1,\ x+e_2} = S_{rs}$ = if $j = 1$, $k = 2$, for example

s_i = ith measurement or reading of I_2

t_o = observed value of t

$t_{1-\alpha}$ = upper α significance level of Student's t, with $\alpha = 0.01, 0.05$, etc., but < 0.5

$t(n-2, A=B)$ = Student's t statistic with $(n-2)$ degrees of freedom for testing hypothesis that $A = B$

$t(n-2, \sigma_x/\sigma_{e_1})$ = Student's t for $(n-2)$ degrees of freedom and a hypothesized value of σ_x/σ_{e_1}. (Applies also to other degrees of freedom and parameters.)

$t(n-2, \sigma_x/\sigma_{e_1}=5)$ = Student's t test based on $(n-2)$ degrees of freedom of the hypothesis that $\sigma_x/\sigma_{e_1} = 5$

t_i = ith measurement or reading of I_3

t_α = upper α probability level of Student's t

$u = r - s$ = difference in readings of instruments I_1 and I_2

u = factor or constant for the upper confidence bound of Hanumara and Thompson (see Eqs. 2-95 and 2-96)

$\bar{u}$ = mean of the difference in readings between instruments I_1 and I_2

$u_i = r_i - s_i = \beta_1 - \beta_2 + e_{i1} - e_{i2}$ = difference in readings of instruments I_1 and I_2 for ith item

Var ( ) = σ^2( ) = population (large sample) variance of the quantity within parentheses

$v = s - t$ = difference in readings of instruments I_2 and I_3

$v_i = s_i - t_i = \beta_2 - \beta_3 + e_{i2} - e_{i3}$ = difference in readings of instruments I_2 and I_3 for the ith item

$w = t - r$ = difference in readings of instruments I_3 and I_1

$w_i = t_i - r_i = \beta_3 - \beta_1 + e_{i3} - e_{i1}$ = difference in readings of instruments I_3 and I_1 for ith item

x = true unknown value of a random variable measured with error

$\bar{x} = \sum_{i=1}^{n} x_i/n$ = sample average of the x_i for n items

x_i = true value of the ith item or characteristic measured

x_{ij} = element or observation in the ith row and jth column of an experimental design

$\bar{z}$ = mean of the readings of instrument I_3 minus the mean of the readings of instruments I_1 plus I_3

α = probability of rejecting the null hypothesis when it is true

α_k = constant in Jaech's model (see Eq. 2-118)

β = true unknown bias or systematic error of a measurement

β_j = constant bias or systematic error of measurement for the jth instrument I_j

β_k = constant in Jaech's model (see Eq. 2-118)

$\delta = 1/k^2$, where $k = \sigma_{e_2}/\sigma_{e_1}$

δ_L = lower $(1-\alpha)$ confidence bound on δ

δ_U = upper $(1-\alpha)$ confidence bound on δ

$\theta = (\sigma_{e_2}^2 + \sigma_{e_3}^2)/(\sigma_{e_1}^2 + \sigma_{e_2}^2)$ = particular ratio of population imprecisions of measurement for three instruments (see Eqs. 2-72 and 2-73)

λ = Wilks' likelihood ratio

λ = likelihood ratio statistic used to test H_0

λ_α = α probability level of the likelihood ratio λ

μ = true unknown (population) value of an item or characteristic measured with error

ν = $[\sigma^2_{e_3} + (\sigma^2_{e_1} + \sigma^2_{e_2})/4]/(\sigma^2_{e_1} + \sigma^2_{e_2})$ = parameter in the t test as in Eq. 2-78

$\rho_{e_1e_2}$ = population correlation coefficient of the errors of I_1 and I_2. (Applies also to any other pair of letters, e.g., *rs, xe, uv,* etc.)

$\sigma(\quad)$ = population standard deviation of quantity in parentheses

σ_e = imprecision standard deviation used when $\sigma_{e_1}=\sigma_{e_2}=\sigma_e$

σ_e = population standard deviation of the errors of measurement

σ^2_{ej} = large sample or population variance of errors of measurement for instrument I_j, $\sigma^2_{e_1}$ being that for I_1, etc.

$\sigma_{e_1e_2}$ = $\rho\sigma_{e_1}\sigma_{e_2}$= large sample or population covariance of the errors of measurement of I_1 and I_2 if it is nonzero

$\text{est}\sigma^2_{e_1}$ = estimate of the population variance of the errors of measurement for instrument I_1

$\text{est}\sigma^2_{e_2}$ = estimate of the population variance of the errors of measurement for instrument I_2

$\text{est}\sigma^2_{e_3}$ = estimate of the population variance of the errors of measurement for instrument I_3

σ_{e_i} = standard deviation of error of calibration at the ith echelon in the hierarchy of calibrations (used in par. 2-11)

σ_L = standard deviation among true laboratory means or levels, or external sigma

σ_{m+1}/σ_m = σ_x/σ_m = precision or "accuracy" ratio in a calibration hierarchy at the last or mth stage

σ_{R_n} = reproducibility sigma = $\sqrt{\sigma^2_L + \sigma^2_r/n}$ for n observations at a laboratory

σ_r = repeatability sigma or standard deviation within laboratories

σ_x = population standard deviation of the true product variability

σ_{xe} = large sample or population covariance of x and e. Indeed, σ_{xe} is the population covariance of the errors of measurement with the level of true values measured and could be estimated by S_{xe}, if isolable.

σ_x/σ_e = product-measurement precision ratio, often misnamed the "accuracy ratio"

$\text{est}\sigma^2_x = \hat{\sigma}^2_x$ = estimate of the unknown population variance σ^2_x

$\chi^2(\quad)$ = chi-square statistic of ( ), the number of degrees of freedom

$\hat{}$ = estimate of quantity under the $\hat{}$

2-1 PRELIMINARY BACKGROUND STATEMENT

A very important and yet widely misunderstood concept or problem in science and technology is that of the precision and accuracy of measurement. It therefore becomes necessary to define errors of measurement and the terms precision and accuracy (or imprecision and inaccuracy) very clearly and then express them in an analytical way. Also we need to present efficient methods of estimating precision and accuracy numerically, and we need to establish or develop appropriate statistical tests of significance for the measures, especially since a relatively small number of measurements usually will be made or taken in most experimental investigations.

In this chapter we will attempt to approach this important problem in a systematic manner and reference some of the key pertinent literature on the subject. In particular, we will (1) give an account of the procedures for estimating the variances in errors of measurement, or the "imprecisions" of measurement, showing that at least two instruments are needed to estimate instrumental imprecisions, and (2) proceed to present techniques for comparing precision of measurement as well as making some useful statements about accuracy and what might be done about it. We believe that most readers will acquire competence in applying the needed techniques if we present illustrative examples as necessary; accordingly, this will be our approach.

The subject matter of this chapter is covered first in the handbook because the statistician analyzes observational data, and the capability of the measurement process should be assessed beforehand.

2-2 INTRODUCTION AND CONCEPT FORMULATION

Each and every measurement or observation can be considered to consist of two "inseparable" components: one is the true value of the item or characteristic being measured, and the other is an error of measurement (instrumental error). The error of measurement of a quantity is widely known as the difference between the observed measurement and the true value of the magnitude of this quantity. The error of measurement is taken to be positive or negative accordingly as the measurement is more or less than the true value. We say "inseparable" because for a single measurement, or a series of measurements from a single measuring instrument, it is not possible to distinguish exactly the size of the true value(s) of the item(s) gaged and the associated error(s) of measurement that is (are) certain to be made. However, as simply as we have stated this premise, we readily encounter some rather important problems or concepts that require clearing up in our description of the two components of the (total) measurement as defined here. First, there is the "true" value of the item or characteristic, which is part of the measurement taken; the "true" value is of primary interest to the user. This "true" value is something that is rarely attained, except perhaps accidentally, for it deals with the concept of "absolute accuracy", so to speak, and may involve many, many measurements or observations to average out the errors committed in the measuring process.

Measurements are an essential part of our daily life, and it is through them that we communicate and make progress in specifying just what is desired, needed, or will be accepted. Thus there must be some basic agreements on just how "accurate" or "true" values will be obtained or sought out, whether they relate to weight or mass, length, time, area, volume, or whatever characteristic is of interest. In any event, the true or "absolute" values of measured items must be made relative to agreed upon standards and methods of measurement. The method of measurement selected should consist of a set of instructions specifying the apparatus and auxiliary equipment to be used to take the observations, the operations to be performed, the sequence in which they are to be carried out, and the conditions under which they are to be respectively taken (Ref. 1, pp. 21-165). Indeed, this is why we have a National Bureau of Standards, which must establish approved methods for measuring and even rule authoritatively on measurements, especially in the event of disagreements. Moreover, and as we shall see, the "perfectly acceptable" measurements will also have to be "precise". But this brings up another important term—accuracy. In this very limited account we have immediately run into two, so far vague, terms that need clarification; namely, "precision" and "accuracy". Accordingly, we must define them, perhaps best in analytical terms, as we proceed and indicate just how they may be quantified and estimated. We return briefly to the concept of true value before proceeding further.

If there were no errors of measurement committed, we would determine the true value of the item being measured each time a measurement is taken. However, in the presence of errors of measurement, which is practically always the case, we have to hypothesize and deal with the more practical situation as described previously. Therefore, it might be helpful if we now consider the concept of a "limiting value". If repeated measurements of a quantity or characteristic are taken and each time the mean of them is calculated, we find that as the number of measurements increases without bound, our calculated means will approach a limiting value. Hence if we were to continue taking such measurements indefinitely and calculating the average of them, we would eventually arrive at a mean value, to some specified or preset number of decimal places, which would not change. The "ultimate" mean value, attained as the number of measurements increases beyond bounds, may be referred to as a limiting value. Unfortunately, this limiting value may not equal exactly the true value of the item measured because on the average there may be some "bias" in the instrument used for measuring or, put otherwise, our measuring instrument has a "systematic error" since the mean of the readings does not approach the true (yet most often unknown) value. Some further quantification of these statements is necessary.

Let us fix the ideas just expressed a little more concretely through the use of a simple, yet appropriate, analytical model. Thus we might well express a single measurement taken with an instrument as

$$r = \mu + e' \tag{2-1}$$

where

r = value of the measurement or the observation itself
μ = true but unknown value of the item measured
e' = error of measurement or the instrumental error.

As an example, one might find that the observed or measured muzzle velocity (MV) of a round fired from a gun or cannon is 659.5 m/s. However, he does not know the *true* MV μ of the projectile nor does he know the size of the error of measurement e' because only the sum of the two components is observed.

As some further introduction, note that in Eq. 2-1 we have used the Greek letter μ for the true unknown or "population" value and the letter e' as the random error of measurement. Had the true value been a random variable, we would have specified it by using the letter x, for example, in the place of μ. The measurement then would have been given as

$$r = x + e' \tag{2-2}$$

where

x = true but unknown random value measured with error.

There is no evidence of any bias or systematic error in either Eq. 2-1 or 2-2 unless the average of a series of measurements is such that the average error of measurement $\bar{e}'$

$$\bar{e}' = \sum_{i=1}^{n} e_i'/n, \tag{2-3}$$

where

n = number of measurements or sample size,

does not approach zero as the number of measurements increases without limit. (The limiting value of the average error would not approach zero.) Thus the large sample average of the errors, or the limiting value, must approach some quantity $\beta \neq 0$ for there to be a bias or systematic error of size β. In this case, we may as well hypothesize that generally the observed measurement should be described as

$$r = \mu + \beta + e \tag{2-4}$$

where

β = instrumental bias or systematic error
e = random error of measurement whose mean or expected value is zero

and the true mean μ (or x) has not changed. We now perceive that for an appropriate general formulation of the measurement problem, we need to hypothesize that any measured value or observation may consist of three inseparable components—first, the true value desired; second, an instrumental bias; and third, a random error of measurement. The total error of measurement consists of the bias error plus the random measurement error, i.e., the sum $(\beta + e)$.

Perhaps the bias β may not normally vary during a series of measurements although by definition we do expect the accidental errors e to be randomly distributed and average out to zero. It is the variation in e that will be used to define and describe the *precision*—or the imprecision—of measurement, and the total error $(\beta + e)$ committed will be used to define and describe the *accuracy* of measurement.

With even this brief formulation of principles, it may be easy for the reader to understand why there is so much confusion about the terms precision and accuracy. The problem becomes very involved because the three components—μ, β, and e—are confounded or inseparable. Indeed, this alone is enough to substantiate that even very intelligent discussions on precision and accuracy may be difficult or somewhat incomprehensible; therefore, we need to proceed cautiously. We will accomplish this by discussing, in appropriate detail, the case of measurements with a single instrument so that our concepts and ideas will be further illuminated. Also we urge the interested reader to study the compendium of papers in Ref. 1 for further background and to read the references and bibliography for further enlightenment.

2-3 MEASUREMENTS WITH A SINGLE INSTRUMENT*

As discussed in par. 2-2, if we were to measure repeatedly the same item or characteristic, the average of a large number of instrumental readings would, according to the model of Eq. 2-4, approach the true value μ plus the inseparable bias β of the measuring instrument if it exists since, under the assumptions used, the average of the errors e would be zero. Hence if this were the applicable model, then for a perfectly calibrated measuring instrument we would not have any great problem with imprecision of measurement for a large number of instrument readings—for example, the determination of the single value of a fundamental physical constant, such as the velocity of light. On the other hand, we must perceive also of the prevalent case, or hypothesize, that the true values may vary from one measurement to another in either a systematic or a random manner. Therefore, a somewhat more appropriate model is of the form $x + \beta + e$, where both x and e are variables, and only the quantity β may be constant over some series of measurements. As an example, consider the series of powder train fuze burning times listed in Table 2-1. These 30 individual burning times are fairly random and illustrate the points we bring out.

TABLE 2-1

BURNING TIMES OF 30 POWDER TRAIN FUZES, s

10.10	9.62	9.50
9.98	10.24	9.56
9.89	9.84	9.54
9.79	9.62	9.89
9.67	9.60	9.53
9.89	9.74	9.52
9.82	10.32	9.44
9.59	9.86	9.67
9.76	10.01	9.77
9.93	9.65	9.86

The average $\bar{r}$ of these $n = 30$ sample values or observations is

$$\bar{r} = \sum_{i=1}^{n} r_i/n \qquad (2\text{-}5)$$

$$= \sum_{i=1}^{30} r_i/30 = 9.7733 \text{ s}$$

where

r_i = ith reading or measurement.

Under the hypothesis that

$$r_i = x_i + \beta + e_i \qquad (2\text{-}6)$$

where

x_i = true value of ith fuze burning time
β = constant instrumental bias if it exists
e_i = random error of measurement for the ith reading

we see that

$$\bar{r} = (1/n)\sum_{i=1}^{n} x_i + \beta + (1/n)\sum_{i=1}^{n} e_i = \bar{x} + \beta + \bar{e} = 9.7733 \text{ s} \qquad (2\text{-}7)$$

*For our purposes, the terms instrument and measurement process may be used interchangeably here.

where

$\bar{x} = \Sigma x_i/n$ = sample average of the x_i for n measurements

$\bar{e} = \Sigma e_i/n$ = sample average of the e_i for n measurement errors.

However, there is absolutely no way to break down the average of 9.7733 s into the three inseparable components of true average fuze burning time $\bar{x}$, the instrumental bias β, and the average error of measurement $\bar{e}$. Thus we are "stuck", as it were, with measurements from a single instrument although we could and should have had our measuring instrument, in this case an electrical clock, calibrated properly before the burning times were taken.

Let us next calculate the sample variance of the 30 fuze times based on $(n - 1) = 29$ degrees of freedom (df). In this connection we define

$$A_{rr} = n \sum_{i=1}^{n} r_i^2 - \left(\sum_{i=1}^{n} r_i\right)^2 \tag{2-8}$$

and see that the sample variance S_r^2 for the data of Table 2-1 is

$$S_r^2 = \sum_{i=1}^{n} (r_i - \bar{r})^2 / (n - 1) = A_{rr}/[n(n - 1)] = 0.04714 \tag{2-9}$$

and the sample standard deviation is $S_r = 0.2171$ s.

If Eq. 2-6 is substituted into Eq. 2-9, we have symbolically

$$S_r^2 = S_x^2 + 2S_{xe} + S_e^2 \tag{2-10}$$

where

$$S_x^2 = \left(\frac{1}{n-1}\right) \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{2-11}$$

$=$ sample variance of the true fuze times

$$S_e^2 = \left(\frac{1}{n-1}\right) \sum_{i=1}^{n} (e_i - \bar{e})^2 \tag{2-12}$$

$=$ sample variance of the errors of measurement

$$S_{xe} = \left(\frac{1}{n-1}\right) \sum_{i=1}^{n} (x_i - \bar{x})\,(e_i - \bar{e}) \tag{2-13}$$

$=$ sample covariance of the true values and the errors of measurement.

Nevertheless, there is no way to decompose properly the variance $S_r^2 = 0.04714$ into the product true variability or sample variance S_x^2 of true fuze times, the variance in errors of measurement or "imprecision" S_e^2, and the covariance between fuze times and errors of measurement S_{xe} since they are confounded. The reader may observe, however, that S_x^2, or its square root S_x, is a measure of the true variability in fuze times; S_e^2, or S_e, is a measure of the dispersion in errors of measurement for the electric clock and the person who operated it, and S_{xe} is a measurement of the "dependence" between the true fuze times and the errors of measurement.

The sample correlation coefficient r_{xe} between true fuze times and errors of measurement would be given by

$$r_{xe} = S_{xe}/(S_x S_e) \tag{2-14}$$

if it could be calculated!

Summarizing, we find that the average $\bar{x}$ of the true values, the bias or systematic error β, and the average $\bar{e}$ of the random errors of measurement are confounded as are the individual values as shown in Eq. 2-6. Also we see that, with proper calibration of the instrument against an authoritative standard, we might be able to reduce the bias of the instrument to near zero or even to zero. Moreover, it can be seen

from Eq. 2-7 that once the bias is elimina ted, and for a large number of measurements and the assumption that the errors of measurement e_i are randomly distributed with zero mean, it is clearly possible to obtain accurately the average $\bar{x}$ of the true values. In acldition, if we are concerned with the determination of a single true value μ, for example, the velocity of light, then from Eq. 2-4 we may approach that value quite closely for an ever increasing number of measurements with a properly calibrated instrument that would not have a systematic error or bias. So much for average values, we must now turn to descriptions of the dispersion or variation in errors of measurement and of the true values themselves.

Taking a close look at the variance S_r^2 of the observations or the measurements as in Eq. 2-10, we see that it also consists of three confounded components. The first or S_x^2 is an efficient measure of the product variability or the variation in the true values of the items measured. Hence S_x^2 is the "product variance", and the square root of it S_x is the standard deviation of the product variability—obviously, a very important component of interest to estimate. Further, the quantity S_e^2 is the sample variance of the errors of measurement and is an excellent representation of the "precision" or the "imprecision" of measurement. Thus if S_e^2 is small, the measurements are considered to be precise; if it is large, the measurements are imprecise. Therefore, we will use this variance S_e^2 of the errors of measurement, or the square root of it S_e, which is the standard error of measurement, to describe the imprecision of measurement. Moreover, the reader may see rather easily that the size of S_e relative to that of S_x would be of considerable importance in the efficiency of most measurement analyses. One notes, incidentally, that if S_e were near zero, or perhaps actually equal to zero, the measurements would be very precise indeed, and, to assure accuracy, he would only have to be concerned with the bias of the instrument—generally, a rather desirable situation. (The reader should note that the constant bias or systematic error β does not appear at all in the calculation of any of the variances, i.e., Eqs. 2-9 through 2-12, since it "cancels out" in the differences of the calculations.)

Finally, the sample covariance term or S_{xe} gives a measure of the "dependence" or "correlation" between the sizes of the true values x_i and the errors of measurement e_i if they happen to be so related. In spite of the well-known fact that large measurements often tend to have large errors of measurement, there exist a large number of situations for which no such correlation or dependence is present, and we may indeed hypothesize that S_{xe} tends to zero—a very plausible assumption for many applications.

The large sample or "expected" value of S_e will approach the true unknown or population value of the standard error of measurement, and we will refer to this limiting value as σ_e. Similarly, the large sample or expected value of S_x will tend toward the true product variability, which we will designate as σ_x—another "population" value, so to speak. We see, therefore, that in approaching the problem of precision and accuracy properly we will need to separate out the sample quantity S_e as the measure of precision (or imprecision), which in turn is an estimator of σ_e. In a like manner, we will need to determine and use S_x as the estimate of true product variability σ_x. We observe that the concept of precision of measurement is not so difficult to understand because an estimate of the standard error of measurement σ_e gives a quantified value that can be used to describe precision or imprecision. On the other hand, the proper concept of accuracy is much more difficult to grasp with profound appreciation because it involves both the instrumental bias β and the random error of measurement e. An accurate measurement is obtained only when the sum $(\beta + e)$ is small, and this is complicated by the fact that the random error of measurement e as described may vary "too much" and perhaps "hide" the bias β. Indeed, to determine the size of the instrumental bias β or to calibrate an instrument properly, the precision of measurement should be "good", i.e., σ_e should be suitably small, or the average of a large number of measurements must be obtained so that $\sigma_e/\sqrt{n}$ is small. We also see that (1) precise measurements may not be accurate because of the possible existence of too large a bias and (2) an unbiased measurement may not be very accurate, except accidentally, if the precision of measurement is poor, i.e., σ_e is large. The best approach to guarantee the accuracy of measurement, therefore, seems to be that of attaining sufficiently good precision and then determining the bias and correcting for it, or eliminating the bias through proper calibration. Unfor-

tunately, the bias may vary from one occasion to another, so an additional component of variance or instrumental error may have to be considered and assessed. It may be found that different instruments will have different systematic errors or biases; the same may be true of the different laboratories performing measurements. Different systematic errors or biases between instruments or laboratories will introduce some additional components of variability, which need quantification in many applications.

We see from the discussion that the separation of product variability and the standard error of measurement, or imprecision, cannot be accomplished with a single measuring instrument. It is for this reason that we must examine the cases in which two or more measuring instruments are used to take the same (series of) measurements or to measure simultaneously the same series of characteristics or items of interest.

2-4 THE SEPARATION OF PRODUCT VARIABILITY AND IMPRECISION OF MEASUREMENT WITH TWO INSTRUMENTS

2-4.1 BASIC OUTLINE AND APPROACH

We will now consider the case for which two instruments, I_1 and I_2, are used to take simultaneous or the same measurements on a series of n items or characteristics that exhibit product variability. Our aim is to find a means of separating the product variability S_x from the imprecision of measurement S_e, i.e., the standard error of measurement. Thus in this case the observed values or the measurements may be represented symbolically as follows:

Measurements by I_1	Measurements by I_2
r_i	s_i
$r_1 = x_1 + \beta_1 + e_{11}$	$s_1 = x_1 + \beta_2 + e_{12}$
$r_2 = x_2 + \beta_1 + e_{21}$	$s_2 = x_2 + \beta_2 + e_{22}$
.	.
.	.
.	.
$r_i = x_i + \beta_1 + e_{i1}$	$s_i = x_i + \beta_1 + e_{i2}$
.	.
.	.
.	.
$r_n = x_n + \beta_1 + e_{n1}$	$s_n = x_n + \beta_2 + e_{n2}$

where

r_i = ith measurement of the first instrument I_1
s_i = ith measurement of the second instrument I_2
x_i = true (unknown) value of ith item
β_1 = bias or systematic error committed by I_1
β_2 = bias or systematic error committed by I_2
e_{i1} = random error of measurement of I_1 on the ith item
e_{i2} = random error of measurement of I_2 on the ith item.

Note that the difference in readings of I_1 and I_2 for the ith item is

$$r_i - s_i = \beta_1 - \beta_2 + e_{i1} - e_{i2} \tag{2-16}$$

and does not include the true value x_i at all.

With reference to these definitive formulations, the sample mean or average value for the measurements of instrument I_1 is from Table 2-1 (the first column in Table 2-2 is a repeat of Table 2-1).

$$\begin{aligned} \bar{r} &= \bar{x} + \beta_1 + \bar{e}_1 \\ &= 9.7733 \text{ s} \end{aligned} \tag{2-17}$$

and that of instrument I_2 is

$$\begin{aligned} \bar{s} &= \bar{x} + \beta_2 + \bar{e}_2 \\ &= 9.7414 \text{ s} \end{aligned} \tag{2-18}$$

using the 29 observations—since one was lost—of the second column of Table 2-2. The difference between the mean measurements of I_1 and I_2 is therefore

$$\bar{r} - \bar{s} = \beta_1 - \beta_2 + \bar{e}_1 - \bar{e}_2 \tag{2-19}$$

and, under the assumption that the random errors have zero means or expected values, Eq. 2-19 gives a more precise estimate of the difference in biases β_1 and β_2 than Eq. 2-16.

Continuing, we see from the definitions of variances and covariances and from Eq. 2-15 that we may calculate three variances and one covariance for the two instruments I_1 and I_2 and have symbolically that

$$S_r^2 = S_x^2 + 2S_{xe_1} + S_{e_1}^2 \tag{2-20}$$

$$S_s^2 = S_x^2 + 2S_{xe_2} + S_{e_2}^2 \tag{2-21}$$

$$S_{rs} = S_x^2 + S_{xe_1} + S_{xe_2} + S_{e_1e_2} \tag{2-22}$$

$$S_{r-s}^2 = S_{e_1}^2 - 2S_{e_1e_2} + S_{e_2}^2 \tag{2-23}$$

where

S_{xe_1} = covariance of true values and errors of measurement of I_1
S_{xe_2} = covariance of true values and errors of measurement of I_2
$S_{e_1e_2}$ = sample covariance of errors of measurement of I_1 and I_2
S_s^2 = sample variance of instrument I_2 based on $(n - 1)$ df
S_{rs} = covariance of the readings of the first and second instruments I_1 and I_2
S_{r-s}^2 = sample variance of the difference in readings of instruments I_1 and I_2.

However, concerning the four equations or calculations, Eqs. 2-20 through 2-23, we may add Eqs. 2-20 and 2-21 and then subtract Eq. 2-22 twice; the result is identically equal to Eq. 2-23. Hence the four equations are linearly dependent. Consequently, for the two-instrument case we really have only three useful equations but six unknown "inseparable" components to estimate. Our primary interest centers around the estimation of product variability and the imprecisions of measurement of the two instruments —i.e., S_x^2, $S_{e_1}^2$, and $S_{e_2}^2$. Hence by assuming that the true values measured and the instrumental errors are mutually or statistically independent of each other, the expected values of the three covariances will vanish, or approach zero, thereby rendering a feasible solution. In fact, as pointed out by Grubbs (Ref. 2), the covariance S_{rs} between the two instrument readings will then approach the product variance, so that for purposes of estimation we have

$$\text{est}\sigma_x^2 = S_{rs} \tag{2-24}$$

$$= (S_{r+s}^2 - S_{r-s}^2)/4 \quad \text{(Ref. 2).}$$

Furthermore, from Ref. 2

$$\text{est}\sigma_{e_1}^2 = S_r^2 - S_{rs} \tag{2-25}$$

$$= (S_r^2 - S_s^2 + S_{r-s}^2)/2$$

and

$$\text{est}\sigma_{e_2}^2 = S_s^2 - S_{rs} \tag{2-26}$$

$$= (S_s^2 - S_r^2 + S_{r-s}^2)/2$$

where

S_{r+s}^2 = sample variance of the sum of readings of instruments I_1 and I_2
$\text{est}\sigma_x^2$ = estimate of unknown population variance σ_x^2
$\text{est}\sigma_{e_1}^2$ = estimate of population variance of the errors of measurement for instrument I_1
$\text{est}\sigma_{e_2}^2$ = estimate of population variance of the errors of measurement for instrument I_2.

The sample or estimated product variance and the variances in errors of measurement of the two instruments are expected to be positive although we see from Eqs. 2-25 and 2-26 that this requires S_{rs} to be smaller than S_r^2 and S_s^2. Often this is not the case as we will see even for respectable sample sizes.

It is also of some interest to note that if the product variance is zero, i.e., S_x^2 and $\sigma_x^2 = 0$, or the same item is measured n times by I_1 and I_2, one might expect that S_{xe_1} and S_{xe_2} would vanish. Thus he would have to contend only with the estimation of σ_{e_1}, σ_{e_2}, and $\sigma_{e_1e_2}$, the covariance of errors of I_1 and I_2, if it exists. In this connection, moreover, a solution using Eqs. 2-20, 2-21, and either 2-22 or 2-23 is clearly obtainable to estimate σ_{e_1}, σ_{e_2}, and $\sigma_{e_1e_2}$.

If there were no errors of measurement, then it is seen that S_r^2, S_s^2, and S_{rs} all give the correct estimate of product variance σ_x^2.

Example 2-1:

We will illustrate the estimation of product variability and imprecision of measurement for the case of two instruments by referring to the data of Table 2-2. The data given there refer to an old, widely analyzed example that appeared in 1948. Nevertheless, it is very useful for our exposition of the applications and problems encountered in the area of estimation of precision of measurement. In Table 2-2 the individual burning times of powder train fuzes are listed as measured by each of three observers on 30 rounds of artillery ammunition fired from a gun. The fuzes were all set for a burning time of 10 s. The "burning time" was defined as the elapsed interval of time from the instant the projectile departed the gun muzzle to the instant of fuze functioning as noted by the flash of the detonating high explosive (at night). The times listed were measured by three electric clocks, each of which was started by a gun muzzle switch, and each clock was stopped independently by an observer as he noticed the flash. We have chosen this particular example because it represents a respectable sample size; nevertheless, it presents some problems relative to the often discouraging occurrence of negative estimates of variance or dispersion, at least for two instruments. For a two-instrument example we will use the measured values r and s of instruments I_1 and I_2, the first two columns, and the differences (4th column). We calculate

$S_r^2 = 0.04714023$ based on all 30 readings of I_1

$S_r^2 = 0.04675448$ based on 29 readings of I_1, excluding 10.01, for which I_2 lost the round

$S_s^2 = 0.045112315$ for $n = 29$ by Eq. 2-12, $S_{rs} = 0.045581897$ for $n = 29$ by Eq. 2-13.

TABLE 2-2

FUZE BURNING TIMES AND DIFFERENCES IN SECONDS

Observer I_1	Observer I_2	Observer I_3	Differences		
r	s	t	$r-s$	$s-t$	$r-t$
10.10	10.07	10.07	0.03	0.00	0.03
9.98	9.90	9.90	0.08	0.00	0.08
9.89	9.85	9.86	0.04	−0.01	0.03
9.79	9.71	9.70	0.08	0.01	0.09
9.67	9.65	9.65	0.02	0.00	0.02
9.89	9.83	9.83	0.06	0.00	0.06
9.82	9.75	9.79	0.07	−0.04	0.03
9.59	9.56	9.59	0.03	−0.03	0.00
9.76	9.68	9.72	0.08	−0.04	0.04
9.93	9.89	9.92	0.04	−0.03	0.01
9.62	9.61	9.64	0.01	−0.03	−0.02
10.24	10.23	10.24	0.01	−0.01	0.00
9.84	9.83	9.86	0.01	−0.03	−0.02
9.62	9.58	9.63	0.04	−0.05	−0.01
9.60	9.60	9.65	0.00	−0.05	−0.05
9.74	9.73	9.74	0.01	−0.01	0.00
10.32	10.32	10.34	0.00	−0.02	−0.02
9.86	9.86	9.86	0.00	0.00	0.00
10.01	lost	10.03	—	—	−0.02
9.65	9.64	9.65	0.01	−0.01	0.00
9.50	9.49	9.50	0.01	−0.01	0.00
9.56	9.56	9.55	0.00	0.01	0.01
9.54	9.53	9.54	0.01	−0.01	0.00
9.89	9.89	9.88	0.00	0.01	0.01
9.53	9.52	9.51	0.01	0.01	0.02
9.52	9.52	9.53	0.00	−0.01	−0.01
9.44	9.43	9.45	0.01	−0.02	−0.01
9.67	9.67	9.67	0.00	0.00	0.00
9.77	9.76	9.78	0.01	−0.02	−0.01
9.86	9.84	9.86	0.02	−0.02	0.00

Consequently, we estimate

$$\text{est}\sigma_x^2 = S_{rs} = 0.04558$$

$$\text{est}\sigma_x = 0.2135 \text{ s}$$

$$\text{est}\sigma_{e_1}^2 = S_r^2 - S_{rs} = 0.001558$$

$$\text{est}\sigma_{e_1} = 0.03947 \ (n = 30)$$

$$\text{est}\sigma_{e_1} = 0.03424 \ (n = 29)$$

$$\text{est}\sigma_{e_2}^2 = S_s^2 - S_{rs} = -0.0004696 < 0, \text{ a slightly negative variance.}$$

Thus even for this large a sample for the two-instrument case, we get a negative variance; therefore, we must take $\sigma_{e_2} = 0$. Negative variances may occur because of random sampling fluctuations (or small sample size, which hardly seems plausible here) or because of a violation of the assumptions, such as the

existence of correlations, or perhaps one or more "outliers". (We cover the analysis of outliers in Chapter 3.) Referring to the data of Table 2-2 and especially the columns of differences, we see that I_1 generally lags I_2 (4th column) except toward the latter rounds and that I_1 is somewhat "ragged". In fact, the mean value of the differences in the fourth column is 0.02379, and the standard error of these differences is 0.02651, as we will see later. Approximate 95% confidence limits on an individual difference may be estimated from $0.02379 \pm 1.96\,(0.02651)$, which gives an interval from about -0.03 to 0.08, so that there are three values (of 0.08) on the upper limit that give the suspicion of poor or ragged times or a lack of good control for I_1.

2-4.2 TREATMENT OF NEGATIVE OBSERVED VARIANCES

There has been much study of the problem of negative estimates of components of variance. This work is beyond the scope of this handbook, and it seems unnecessary to delve into the subject extensively here. However, it is of some interest to point out that Thompson (Ref. 3), working with a method of modified maximum likelihood estimation, has suggested treating negative variance estimates in accordance with the rules given in Table 2-3.

TABLE 2-3

NONNEGATIVE VARIANCE ESTIMATES
THE TWO-INSTRUMENT CASE (Ref. 3)*

If	Take $\text{est}\sigma_x^2 =$	Take $\text{est}\sigma_{e_1}^2 =$	Take $\text{est}\sigma_{e_2}^2 =$
$S_r^2 > S_{rs}$ $S_s^2 > S_{rs} > 0$	S_{rs}	$S_r^2 - S_{rs}$	$S_s^2 - S_{rs}$
$S_r^2 > S_{rs} > S_s^2$	S_s^2	$S_r^2 + S_s^2 - 2S_{rs}$	0
$S_s^2 > S_{rs} > S_r^2$	S_r^2	0	$S_r^2 + S_s^2 - 2S_{rs}$
$S_{rs} < 0$	0	S_r^2	S_s^2

Reprinted with permission. Copyright © by American Statistical Association.

For our application, therefore, we would, according to Thompson (Ref. 3), take

$$\text{est}\sigma_x^2 = S_s^2 = 0.04511 \qquad \text{(the smallest variance)}$$

$$\text{est}\sigma_{e_1}^2 = S_r^2 + S_s^2 - 2S_{rs} = 0.001089 \qquad (n = 30)$$

$$\text{est}\sigma_{e_2}^2 = 0.$$

This decreases $\text{est}\sigma_{e_1}$ from 0.03947 to 0.03299, whereas $\text{est}\sigma_x$ changes from 0.2135 to 0.2124, and $\text{est}\sigma_{e_2}$ has to be taken as zero anyway.

In addition to Thompson's modified ML method of treatment and the possibility that small sample size or the existence of outliers might cause negative estimates of variance, we should also consider the possibility that some of the covariances are real—i.e., that perhaps the errors of measurement are correlated

* In Ref. 4, Hanumara proposes some nonnegative estimates of imprecisions of measurement for the three-instrument case. In par. 2-5 we give in some detail the maximum likelihood (ML) estimates which are ordinarily recommended for use in applications.

with each other or are possibly correlated with the level of true values measured. Of course, there is "quite a game" concerning just what the best or true hypothesis might be in the absence of appropriate information, and one might well have to examine his particular set of data closely to make a valid judgment. If the variation of true values is not over a wide interval, it could be hypothesized that the errors of measurement are correlated. This particular problem has recently been studied by Yang (Ref. 5). Yang's treatment assumes that S_r^2 is the largest variance and estimates $\sigma_x^2 + \sigma_{e_1}^2$ and that S_s^2 estimates $\sigma_x^2 + \sigma_{e_2}^2$ as before, but that due to correlated errors, S_{rs} would estimate the population values given by

$$E(S_{rs}) = \sigma_x^2 + \sigma_{e_1 e_2} = \sigma_x^2 + \rho\sigma_{e_1}\sigma_{e_2} \tag{2-27}$$

where

ρ = true unknown population correlation coefficient of I_1 and I_2 errors

$\sigma_{e_1 e_2}$ = large sample or population covariance of the errors of I_1 and I_2 if it is nonzero.

This approach therefore brings forth the need to treat and estimate another unknown ρ, if it exists, for the data under study. In this connection, one also notes that the large sample or expected values of Eqs. 2-25 and 2-26 then become

$$E(S_r^2 - S_{rs}) = \sigma_{e_1}^2 - \rho\sigma_{e_1}\sigma_{e_2} \tag{2-28}$$

and

$$E(S_s^2 - S_{rs}) = \sigma_{e_2}^2 - \rho\sigma_{e_1}\sigma_{e_2} \tag{2-29}$$

$E(S_r^2 - S_{rs})$ = expected value of the estimate of the population variance of errors of measurement for instrument I_1 if the covariance of errors is zero

$E(S_s^2 - S_{rs})$ = expected value of the estimate of the population variance of errors of measurement for instrument I_2 if the covariance of errors is zero.

Yang (Ref. 5) suggests that the lower bound of the unknown ρ may be estimated from

$$1 \geq \rho^2 \geq -4(S_r^2 - S_{rs})(S_s^2 - S_{rs})/(S_r^2 - S_s^2)^2 \tag{2-30}$$

where we have also indicated that the upper bound of ρ^2 has to be unity, of course. Ref. 5 also suggests the use of the lower bound given by Eq. 2-30 if $|S_r^2 - S_s^2|/(S_r^2 - S_{rs})$ is "close to unity"; if not, the midpoint of the extreme values of Eq. 2-30 should be used, i.e., take

$$\rho^2 \approx (1/2)\,(1 + \text{RHS of Eq. 2-30}) \tag{2-31}$$

where

RHS = "right-hand side of".

This means that putting

$$K = [(S_r^2 - S_s^2)^2 - 4(S_r^2 - S_{rs})(S_s^2 - S_{rs})]^{1/2}. \tag{2-32}$$

Then $\sigma_{e_1}^2$ and $\sigma_{e_2}^2$ are to be estimated from

$$\text{est}\sigma_{e_1}^2 = (S_r^2 - S_s^2)(3S_r^2 - 2S_{rs} - S_s^2 \pm K)/[2(S_r^2 - 2S_{rs} + S_s^2)] \tag{2-33}$$

$$\text{est}\sigma_{e_2}^2 = (S_s^2 - S_r^2)(3S_s^2 - 2S_{rs} - S_r^2 \mp K)/[2(S_r^2 - 2S_{rs} + S_s^2)]. \tag{2-34}$$

The upper signs before K—i.e., + in Eq. 2-33 and − in Eq. 2-34—are to be used if $|S_r^2 - S_s^2|/(S_r^2 - S_{rs})$ is very close to unity (Ref. 5), and the lower signs before K, i.e., − and +, otherwise.

The estimate of product variance σ_x^2 is then found to be

$$\text{est}\sigma_x^2 = S_r^2 - \text{est}\sigma_{e_1}^2 = S_s^2 - \text{est}\sigma_{e_2}^2. \tag{2-35}$$

where $\text{est}\sigma_{e_1}^2$ and $\text{est}\sigma_{e_2}^2$ are calculated, using Eqs. 2-33 and 2-34, respectively.

Using the data of Example 2-1, we find from Eq. 2-30 that Yang's estimated lower bound for ρ^2 is

$$\rho^2 \geq 0.7118$$

and

$$|S_s^2 - S_{rs}|/(S_r^2 - S_{rs}) = 0.3013 \text{ (assumes } n = 30 \text{ for } S_r^2)$$

is not close to unity; accordingly, the lower signs before K in Eqs. 2-33 and 2-34 should be used. By doing so, we obtain

$$\text{est}\sigma_{e_1} \approx 0.04817$$
$$\text{est}\sigma_{e_2} \approx 0.01710^*$$

and from Eq. 2-35

$$\text{est}\sigma_x \approx (0.04714 - 0.002320)^{1/2} = 0.2117$$

as contrasted to 0.2135 determined before.

In summary, we see that Yang's estimators have the desirable property of being both nonnegative and nonzero; however, we will see that his imprecision estimates are high as judged by the more precise case where all three instruments are used (par. 2-5). Moreover, we accomplish an additional advantage by simultaneously using three measuring instruments as in par. 2-5—as indicated by I_1, I_2, and I_3 in Table 2-2—this case being formulated to use only the differences in instrumental errors of measurement, completely free of product true values.

With these attempts, and even for the respectable sample size of 29 or 30, we see that the two-instrument case may lead to somewhat disappointing results although the negative estimates of variance need not bother us too much. Indeed, for any very important experiment of measurement, it may be well to employ three or more instruments, or laboratories, or alternatively we can always use a very satisfying statistical test of significance for the two-instrument case; this test is discussed next.

2-4.3 A SIGNIFICANCE TEST ON IMPRECISION BASED ON TWO INSTRUMENTS

Fortunately, we need not be too concerned by occasional, or even frequent, negative estimates of variance for instrument imprecision. This is because a significance test is available concerning a hypothesized ratio of the product standard deviation to the standard error of measurement. This statistical test of significance was developed by Thompson (Ref. 3), who based it on a result of Roy and Bose's (Ref. 6). The procedure consists of specifying the ratio σ_x/σ_{e_1}** (or σ_x/σ_{e_2}) as a measure of relative precision in which one might be primarily interested and then making a Student's t test to see whether the test would reject the null hypothesis concerning that ratio. In other words, if $\sigma_x/\sigma_{e_1} = 5$ is acceptable, which indicates that the standard error of measurement is only one-fifth that of product variability or true value standard deviation, the precision of measurement is quite satisfactory. On the other hand, if for example the ratio were as small as $\sigma_x/\sigma_{e_1} = 1$ or even 2, the relative precision of measurement would be so poor that a more precise measuring instrument would be required. The Student's t test suggested by Thompson (Ref. 3) is, using $(n - 2)$ df,

$$t(n-2, \sigma_x/\sigma_{e_1}) = \sqrt{n-2}\,[S_r^4/(S_r^2S_s^2 - S_{rs}^2)]^{1/2}\ [(S_{rs}/S_r^2) - \sigma_x^2/\ (\sigma_x^2 + \sigma_{e_1}^2)]. \tag{2-36}$$

*Some recent results have been obtained. See Ref. 5.

**This ratio is often referred to as the "accuracy ratio" although the term product/precision of measurement ratio or simply precision ratio would be much better.

By taking $t_{1-\alpha}$ equal to the upper α probability level or percentage point of the Student's t distribution, Eq. 2-36 is less than $t_{1-\alpha}$ if and only if

$$\sigma_x^2/\sigma_{e_1}^2 > \frac{S_{rs} - t_{1-\alpha}[(S_r^2 S_s^2 - S_{rs}^2)/(n-2)]^{1/2}}{S_r^2 - S_{rs} + t_{1-\alpha}[(S_r^2 S_s^2 - S_{rs}^2)/(n-2)]^{1/2}} \quad * \tag{2-37}$$

A very similar test for σ_x/σ_{e_2} relative to the second instrument is readily obtained by replacing the first S_r^2 in the denominator of Eq. 2-37 with S_s^2, or similarly S_r^4 by S_s^4, and S_{rs}/S_r^2 by S_{rs}/S_s^2 in Eq. 2-36.

Example 2-2:

Referring to Example 2-1, we are not concerned about the imprecision of measurement for I_2 because of the near zero standard error of measurement, but let us test the hypothesis that $\sigma_x/\sigma_{e_1} = 5$ at the upper 5% level.

By using Eq. 2-36, we calculate for $n = 29$ readings for I_1

$$t(27, \sigma_x/\sigma_{e_1} = 5) = \sqrt{27}\left[\frac{(0.04675)^2}{(0.04675)\,(0.04511) - (0.04558)^2}\right]^{1/2}\left(\frac{0.04558}{0.04675} - \frac{25}{26}\right) = 0.583$$

whereas $t_{0.95}(27) = 1.703$. Hence we accept the null hypothesis that $\sigma_x/\sigma_{e_1} \geq 5$ for our measurement process. We note in passing that if we stated $\sigma_x/\sigma_{e_1} = 3.82$, this hypothesis would be just barely rejectable at $Pr = 0.95$.

Actually, an estimate of $\sigma_e = 0.03$ or 0.04 for either measuring instrument may not be very good for estimating the true value of burning time for a *single* round although for the *average* of 30 rounds, the value of $\sigma_e/\sqrt{30} = 0.04/\sqrt{30} = 0.007$ may not be considered too poor. Finally, concerning true product variability, we see that

$$\sqrt{S_r^2} = \sqrt{0.04714} = 0.2171 \text{ s} \quad (n = 30)$$

and

$$\sqrt{S_{rs}} = \text{est}\sigma_x = 0.2135 \text{ s} \quad (n = 29)$$

which perhaps shows a small or negligible difference for the effect of σ_{e_1} on the true variability of the product.

2-4.4 VARIANCES OF ESTIMATORS OF IMPRECISION OF I_1 AND I_2

For many applications it is often proper to assume that the product values x_i and the errors of measurement e are normally distributed or approximately so. For this case and the use of two instruments, Grubbs (Ref. 2) derived variances of the estimators—Eqs. 2-24, 2-25, and 2-26—in 1948 to obtain some idea of the reliability or precision and stability of results. As given in Ref. 2, the population variance of the estimate of $\sigma_{e_1}^2$ is

$$\text{Var}(\text{est}\sigma_{e_1}^2) = E(\text{est}(\sigma_{e_1}^2) - \sigma_{e_1}^2)^2$$

$$= \left(\frac{2}{n-1}\right)\sigma_{e_1}^4 + \left(\frac{1}{n-1}\right)(\sigma_x^2\sigma_{e_1}^2 + \sigma_x^2\sigma_{e_2}^2 + \sigma_{e_1}^2\sigma_{e_2}^2). \tag{2-38}$$

Likewise, the population variance of the estimate of $\sigma_{e_2}^2$ is given by

*For an upper bound, the signs of the $t_{1-\alpha}$'s are reversed.

$$\text{Var}(\text{est}\sigma_{e_2}^2) = E(\text{est}\sigma_{e_2}^2 - \sigma_{e_2}^2)^2$$

$$= \left(\frac{2}{n-1}\right)\sigma_{e_2}^4 + \left(\frac{1}{n-1}\right)(\sigma_x^2\sigma_{e_1}^2 + \sigma_x^2\sigma_{e_2}^2 + \sigma_{e_1}^2\sigma_{e_2}^2) \tag{2-39}$$

and the population variance of the estimate of product variability is given by

$$\text{Var}(\text{est}\sigma_x^2 = E(\text{est}\sigma_x^2 - \sigma_x^2)^2$$

$$= \left(\frac{2}{n-1}\right)\sigma_x^4 + \left(\frac{1}{n-1}\right)(\sigma_x^2\sigma_{e_1}^2 + \sigma_x^2\sigma_{e_2}^2 + \sigma_{e_1}^2\sigma_{e_2}^2). \tag{2-40}$$

It is noted that the $\text{Var}(\text{est}\sigma_{e_1}^2)$ depends on (1) σ_x^2, the variance in the characteristic measured; (2) $\sigma_{e_1}^2$, the variance of the errors of measurement of instrument I_1; (3) $\sigma_{e_2}^2$, the variance of the errors of measurement of instrument I_2; and (4) n, the number of observations or the sample size. Therefore, to obtain a precise estimate of $\sigma_{e_1}^2$ when using only two instruments, the variation in the characteristic measured, i.e., σ_x^2, should be held to a reasonable minimum to study imprecision, or the sample size n should be sufficiently large for two instruments.

If the variation in the characteristic measured is zero (or if we measure the same item repeatedly), i.e., if $\sigma_x^2 = 0$, one could compute

$$\text{est}\sigma_{e_1}^2 = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(e_{i1} - \bar{e}_1)^2 \tag{2-41}$$

directly with the variance of the $\sigma_{e_1}^2$ equal to

$$\text{Var}(\text{est}\sigma_{e_1}^2) = \left(\frac{2}{n-1}\right)\sigma_{e_1}^4. \tag{2-42}$$

Apparently, when employing two instruments, there are only two straightforward computational procedures of interest for separating the variability in the product from the variance in the errors of measurement, and both methods give the same estimate. In using either method, however, it is possible to estimate $\sigma_{e_1}^2$, $\sigma_{e_2}^2$, and σ_x^2 and thus determine from the relative order of magnitude of these quantities whether the instruments are sufficiently precise for use in taking the required measurements.

For the two-instrument case the experimentalist may employ very similar or the same kind of instruments. Let us suppose that this is the case, so that

$$\sigma_{e_1}^2 = \sigma_{e_2}^2 = \sigma_e^2.$$

Then Eq. 2-38 becomes

$$\text{Var}(\text{est}\sigma_{e_1}^2) \text{ or } \text{Var}(\text{est}\sigma_{e_2}^2) = \left(\frac{2}{n-1}\right)\sigma_e^4 + \left(\frac{1}{n-1}\right)(\sigma_e^4 + 2\sigma_x^2\sigma_e^2) \tag{2-43}$$

which also involves product variability σ_x^2.

Although it seems not entirely satisfactory to calculate the reduced Eq. 2-38 or Eq. 2-43 when our estimate of σ_{e_2} is zero, we may get some rough idea of the variance of the estimate of $\sigma_{e_1}^2$ in Example 2-1. It is

$$\text{Var}(\text{est}\sigma^2_{e_1}) \approx (2/29)\,(0.001558)^2 + (1/29)\,[(0.001558)^2 + 2(0.04558)\,(0.001558)] = 0.000005148. \tag{2-44}$$

Thus the standard error of the $\text{est}\sigma^2_{e_1} = 0.002269$, which is larger than the estimate itself!

One is bound to feel somewhat uncomfortable about obtaining the estimate of imprecision of the first instrument I_1 as $\text{est}\sigma^2_{e_1} = 0.001558$ and then finding that the expected standard error of that estimate is even larger. This may be due partly to the fact that the estimated σ^2_x of 0.04558 is 29 times the estimated $\sigma^2_{e_1} = 0.001558$. Expressed another way, the second term of Eq. 2-44 is about 30 times the first, which is free of the product variability σ_x. Hence using three instruments may definitely be of considerable interest and value.

2-5 THE SEPARATION OF PRODUCT VARIABILITY AND INSTRUMENT IMPRECISION WITH THREE INSTRUMENTS

By using three instruments to measure either simultaneously or the same series of items or characteristics and by working with the three sets of differences in readings, the product values cancel out and only the differences in instrument biases and random errors remain. Thus if the errors of measurement are relatively small or if the biases are constant and the variance of random errors is a rather low fraction of product variance, then it would be expected that more precise estimates of the imprecision of measurement would be obtained from three instruments as compared to two.

Let us represent the ith reading of the third instrument I_3 symbolically by

$$t_i = x_i + \beta_3 + e_{i_3}. \tag{2-45}$$

We then have the three differences in instrument readings given by

$$u_i = r_i - s_i = \beta_1 - \beta_2 + e_{i_1} - e_{i_2} \tag{2-46}$$

$$v_i = s_i - t_i = \beta_2 - \beta_3 + e_{i_2} - e_{i_3} \tag{2-47}$$

$$w_i = t_i - r_i = \beta_3 - \beta_1 + e_{i_3} - e_{i_1} \tag{2-48}$$

where

u_i = difference in readings of instruments I_1 and I_2 for the ith item
v_i = difference in readings of instruments I_2 and I_3 for the ith item
w_i = difference in readings of instruments I_3 and I_1 for the ith item.

Eqs. 2-46, 2-47, and 2-48 are completely free of any product or true values and involve only differences in the constant biases and differences in random errors of measurement of the three pairs of instruments. Hence it is easily seen that if the instrumental errors are uncorrelated or are statistically independent, the three instrumental imprecisions may be easily and efficiently estimated. In fact, as shown by Grubbs (Ref. 2), the appropriate estimates of imprecision are

$$\text{est}\sigma^2_{e_1} = (S_u^2 - S_v^2 + S_w^2)/2 = S_r^2 - S_{rs} - S_{rt} + S_{st} \tag{2-49}$$

$$\text{est}\sigma^2_{e_2} = (S_u^2 + S_v^2 - S_w^2)/2 = S_s^2 - S_{rs} + S_{rt} - S_{st} \tag{2-50}$$

$$\text{est}\,\sigma^2_{e_3} = (-S_u^2 + S_v^2 + S_w^2)/2 = S_t^2 + S_{rs} - S_{rt} - S_{st} \tag{2-51}$$

where

S_u^2 = sample variance of the difference in readings of instruments I_1 and I_2
S_v^2 = sample variance of the difference in readings of instruments I_2 and I_3
S_w^2 = sample variance of the difference in readings of instruments I_3 and I_1
S_{rt} = covariance of the readings of instruments I_1 and I_3
S_{st} = covariance of the readings of instruments I_2 and I_3
S_{rs} = covariance of the readings of instruments I_1 and I_2.

Even though the variance and covariance terms of each second-listed RHS involve product true values, the estimates of imprecision for the three-instrument case are entirely free of product level. For example, the second-listed RHS of Eq. 2-49 is symbolically

$$\text{est}\sigma_{e_1}^2 = S_{e_1}^2 - S_{e_1e_2} - S_{e_1e_3} + S_{e_2e_3}. \tag{2-52}$$

It contains no x's.

For independent and normally distributed errors of measurement, the variances of the three estimates of instrument imprecision are (Ref. 2)

$$\text{Var}(\text{est}\sigma_{e_1}^2) = \frac{2}{n-1}\left(\sigma_{e_1}^4\right) + \left(\frac{1}{n-1}\right)(\sigma_{e_1}^2\sigma_{e_2}^2 + \sigma_{e_1}^2\sigma_{e_3}^2 + \sigma_{e_2}^2\sigma_{e_3}^2) \tag{2-53}$$

$$\text{Var}(\text{est}\sigma_{e_2}^2) = \frac{2}{n-1}\left(\sigma_{e_2}^4\right) + \left(\frac{1}{n-1}\right)(\sigma_{e_1}^2\sigma_{e_2}^2 + \sigma_{e_1}^2\sigma_{e_3}^2 + \sigma_{e_2}^2\sigma_{e_3}^2) \tag{2-54}$$

$$\text{Var}(\text{est}\sigma_{e_3}^2) = \frac{2}{n-1}\left(\sigma_{e_3}^4\right) + \left(\frac{1}{n-1}\right)(\sigma_{e_1}^2\sigma_{e_2}^2 + \sigma_{e_1}^2\sigma_{e_3}^2 + \sigma_{e_2}^2\sigma_{e_3}^2). \tag{2-55}$$

Note also that the variances of the estimated variances of errors of measurement are free of product variance σ_x^2 and, correspondingly, should be smaller.

The estimate of product variability or the variance of true values is simply the average of all three covariances of the readings of the three instruments. Thus

$$\begin{aligned}\text{est}\sigma_x^2 &= (S_{rs} + S_{rt} + S_{st})/3 \\ &= \frac{1}{9}\left[S_{r+s+t}^2 - \frac{1}{2}(S_u^2 + S_v^2 + S_w^2)\right] \\ &= S_{\bar{r}+\bar{s}+\bar{t}}^2 - \frac{1}{18}(S_u^2 + S_v^2 + S_w^2)\end{aligned} \tag{2-56}$$

where

S_{r+s+t}^2 = sample variance of the sum of the three instrument readings for each item measured
$S_{\bar{r}+\bar{s}+\bar{t}}^2$ = sample variance of the average of the three instrument readings for each item measured.

The variance of Eq. 2-56 is

$$\begin{aligned}\text{Var}(\text{est}\sigma_x^2) = &\left(\frac{2}{n-1}\right)\sigma_x^4 + \left[\frac{4}{9}(\sigma_x^2\,\sigma_{e_1}^2 + \sigma_x^2\,\sigma_{e_2}^2 + \sigma_x^2\,\sigma_{e_3}^2)\right. \\ &\left. + \frac{1}{9}(\sigma_{e_1}^2\,\sigma_{e_2}^2 + \sigma_{e_1}^2\,\sigma_{e_3}^2 + \sigma_{e_2}^2\,\sigma_{e_3}^2)\right].\end{aligned} \tag{2-57}$$

Example 2-3:

Given: The data of Table 2-2 for three simultaneous instrument readings on fuze burning times for the 30 time fuzes.

Find: The best estimates of instrument imprecisions, the round-to-round true dispersion, and determine the variances and standard errors of the estimates.

Using the last three columns or differences in readings of pairs of instruments on each fuze time, we calculate

$$S_u^2 = S_{r-s}^2 = 0.0007030 \text{ s}^2$$

$$S_v^2 = S_{s-t}^2 = 0.0008878 \text{ s}^2$$

$$S_w^2 = S_{t-r}^2 = 0.0003108 \text{ s}^2.$$

Then from Eqs. 2-49, 2-50, 2-51, and 2-56 we obtain

$$\begin{aligned}
\text{est}\sigma_{e_1}^2 &= (0.0007030 - 0.0008878 + 0.0003108)/2 \\
&= 0.0000630^* \\
\text{est}\sigma_{e_1} &= 0.00794 \text{ s} \\
\\
\text{est}\sigma_{e_2}^2 &= (0.0007030 + 0.0008878 - 0.0003108)/2 \\
&= 0.000640^* \\
\text{est}\sigma_{e_2} &= 0.0253 \text{ s} \\
\\
\text{est}\sigma_{e_3}^2 &= (-0.0007030 + 0.0008878 + 0.0003108)/2 \\
&= 0.0002478^* \\
\text{est}\sigma_{e_3} &= 0.015 \text{ s} \\
\\
\text{est}\sigma_x^2 &= 0.046087 - (1/18)\,(0.0007030 + 0.0008878 + 0.0003108) \\
&= 0.04598^* \\
\text{est}\sigma_x &= 0.2144 \text{ s}.
\end{aligned}$$

We note that all three estimates of instrumental imprecision are always positive; that they are straightforwardly estimated from the difference in errors of measurement without questionable boundary conditions; that instrument I_1 is the more precise one, and that I_2 is the worst of the three. Thus the addition of the third instrument to the case of only the first two, where negative variance estimates were obtained, certainly seems quite worthwhile, or even sorely needed. We do not actually know whether these instrumental errors are correlated or whether the covariance terms otherwise really have nonzero expectation although the estimates of imprecision based on the Yang (Ref. 5) approach for I_1 and I_2 are rather high as we now see.

Using Eqs. 2-53, 2-54, and 2-55 next and the previously determined estimates, we calculate the variances and standard errors of the estimators:

$$\begin{aligned}
\text{Var}(\text{est}\sigma_{e_1}^2) &= 0.00000000767 \\
\sigma(\text{est}\sigma_{e_1}^2) &= 0.0000876
\end{aligned}$$

*For readers interested in a Bayesian approach to the estimation of precision of measurement, see Draper and Guttman (Ref. 7). They obtain $\text{est}(\sigma_{e_1}^2/\sigma_x^2) = 0.010675$, $\text{est}(\sigma_{e_2}^2/\sigma_x^2) = 0.001060$, and $\text{est}(\sigma_{e_3}^2/\sigma_x^2) = 0.004109$, whereas our equivalent estimates of these ratios are 0.00137, 0.0139, and 0.00539, respectively.

$$\text{Var}(\text{est}\sigma_{e_2}^2) = 0.00000003565$$
$$\sigma(\text{est}\sigma_{e_2}^2) = 0.000189$$

$$\text{Var}(\text{est}\sigma_{e_3}^2) = 0.0000000116$$
$$\sigma(\text{est}\sigma_{e_3}^2) = 0.000108$$

where

$\sigma(\quad)$ = population standard deviation of quantity in parentheses.

These values are much smaller than corresponding values for the two-instrument case as would be expected since they are free of product variation. Therefore, the three-instrument estimates are quite worthy of adoption since they are entirely satisfactory and conclusive in nature.

For the product variability we have from Eq. 2-57

$$\text{Var}(\text{est}\sigma_x^2) = 0.000165$$
$$\sigma(\text{est}\sigma_x^2) = 0.0128$$

which is $0.0128/0.0000876 - 146$ times $\sigma(\text{est}\sigma_{e_1}^2)$!

With this example and the informative numerical values or estimates obtained, we begin to see the advantage of employing three or more instruments to study precision and accuracy of measurement. Indeed, the use of three measuring instruments should be considered neither an extravagance nor a luxury, especially since it may take three or more instruments to reduce the variances of the estimates of imprecision to suitable values for precise understanding of instrument capability. In fact, the use of several instruments in any important measurement study leads to the idea of "interlaboratory testing", which has long been practiced by the chemical and other industries for the purpose of quantifying precision and accuracy. Moreover, it has been wide practice to measure standard material at even ten or more laboratories in a "round-robin" procedure—as such studies indicate which laboratories are imprecise and inaccurate as well—so that the offenders may be "brought into line". The standard error of measurement at a single laboratory is often referred to as the "repeatability" sigma, whereas that among the laboratories —which includes the standard error of an average value for a single laboratory—is called the "reproducibility" sigma.

Having given a somewhat extensive account of the estimation problem for two and three instruments, we will now give several important statistical tests of significance concerning precision and accuracy, which supply the most desirable type of information.

2-6 SIGNIFICANCE TESTS FOR PRECISION AND ACCURACY OF TWO INSTRUMENTS

2-6.1 PRELIMINARY COMMENTS ON SIGNIFICANCE TESTS FOR TWO INSTRUMENTS

While the estimation of precision and accuracy of measurement parameters is important, comparisons of the relative values of the unknown parameters are also very essential and may be used as a basis for action. For example, consider the two-instrument case for measurements. Here we would like to compare the unknown precision or imprecision of instrument 1 with that of instrument 2 on the basis of, "Does I_1 have a larger or smaller standard error of measurement than I_2?". If the instruments are of the same type, it would be expected that they would have equal standard errors of measurement although one might be poorer than the other if it is not used properly, has been damaged, etc. Once the question of relative precision of measurement has been answered, it becomes quite important to determine whether there is a difference in constant bias of the two instruments. If a test of significance indicates there is a significant difference in biases or systematic errors, the instruments should be calibrated to read properly.

The test of precision is a test of whether σ_{e_1} is equal to, greater than, or less than σ_{e_2}. Should it be true that one or both of the instruments has too large a standard error of measurement, there may be quite a fundamental problem in correcting the difficulty. On the other hand, it could be satisfactory that an increase in the number of measurements will lead to suitable precision, perhaps especially for the average measured value. Fortunately, from this test one also may settle the problem concerning whether the standard error of measurement of one of the instruments is some specified multiple of that of the other. This will be illustrated in the sequel.

Regardless of whether or not it is possible or economical to reduce standard errors of measurement of the two instruments to suitable values if they are much too large, it is nevertheless of great importance to determine whether calibration is called for or at least to make a correction in the readings of one or even both instruments. The statistical test of significance used in this connection determines whether we can say that the bias β_1 of the first instrument equals the bias β_2 of the second instrument or whether one is larger than the other.

2-6.2 TEST OF WHETHER $\sigma_{e_1} = \sigma_{e_2}$ (PRECISION COMPARISON)

The test on relative precision of measurement involves taking the sum p_i and the differences u_i of the readings of the two instruments, i.e., I_1 and I_2, for example, which are

$$p_i = r_i + s_i = \beta_1 + \beta_2 + 2x_i + e_{i1} + e_{i2} \tag{2-58}$$

$$u_i = r_i - s_i = \beta_1 - \beta_2 + e_{i1} - e_{i2}. \tag{2-59}$$

On the assumption of statistically uncorrelated errors of measurement and true values, it is easy to see that the population or expected correlation coefficient ρ_{pu} of p and u is

$$\rho_{pu} = \frac{\sigma_{e_1}^2 - \sigma_{e_2}^2}{[(4\sigma_x^2 + \sigma_{e_1}^2 + \sigma_{e_2}^2)(\sigma_{e_1}^2 + \sigma_{e_2}^2)]^{1/2}} \tag{2-60}$$

and hence that the test of whether $\sigma_{e_1} = \sigma_{e_2}$ is precisely a test of whether the population correlation $\rho_{pu} = 0$. This is easily accomplished on the basis of the Pitman-Morgan test (Refs. 8 and 9) as developed for the purpose by Maloney and Rastogi (Ref. 10). In this connection, one simply calculates the sample correlation coefficient r_{pu} and refers it to a table of percentage points of the correlation coefficient of the bivariate normal distribution or uses the ordinary Student's t test given by Eq. 2-62. First, the sample correlation coefficient is given by

$$r_{pu} = (S_r^2 - S_s^2)/[(S_r^2 + S_s^2 + 2S_{rs})(S_r^2 + S_s^2 - 2S_{rs})]^{1/2}$$

also

$$r_{pu} = \frac{S_{pu}}{S_p S_u}. \tag{2-61}$$

Then the Student's t test based on $(n-2)$ df is

$$t(n-2, \sigma_{e_1} = \sigma_{e_2}) = r_{pu}(n-2)^{1/2}/(1 - r_{pu}^2)^{1/2}$$

$$= \frac{[(S_r^2/S_s^2) - 1](n-2)^{1/2}}{[4(1 - r_{rs}^2) S_r^2/S_s^2]^{1/2}}. \tag{2-62}$$

We will illustrate this test with an example (Example 2-4) of O'Bryon (Ref. 11) concerning the precision and accuracy of velocity chronographs. Also we thought it desirable to illustrate calculations for a smaller sample size, and hence less stable results, than for the data of Table 2-2. This problem arose from a NATO study on velocity chronographs submitted for acceptance or standardization. It was apparently desirable to use two reference or "standard" chronographs, since two are better than one reference instrument, to judge a third chronograph submitted for acceptance. Perhaps it was considered that such a procedure would result in more confidence and provide some checks on the test results. The choice of the two standards for initial tests is somewhat arbitrary indeed although pair wise comparisons of the three instruments can be made simply by permuting the instrument designations—i.e., the r_i, s_i, and t_i—as desired. We examine I_1 and I_2 only at this point.

Example 2-4:

Three velocity-measuring chronographs, the "Fotobalk", the "Counter", and the "Terma" instruments, were used simultaneously to determine velocities of each of twelve successive rounds fired from a 155-mm howitzer*. The velocities were recorded in meters per second (m/s), and the individual velocity measurements are given in Table 2-4. Also recorded in Table 2-4 are the sample variances, the estimated imprecisions of measurement, the estimated differences in biases or systematic errors, and estimated true product variability. We assume here that no past data are available on precision of measurement for the "standard" instruments, the Fotobalk and the Counter, and our purpose ultimately is to check out the precision and accuracy of measurement for the Terma, or "test", instrument. Eqs. 2-49 through 2-51 are used to estimate the standard deviations in errors of measurement for each of the three instruments; the computations are shown in Table 2-4. The estimated standard error of measurement (0.468 m/s) for the Terma chronograph seems larger than that for the other two chronographs. We will check this value later after checking out the two "standards", the Fotobalk and Counter—designated I_1 and I_2—for relative precision and agreement in level of measurement or for bias.

First, we find the sums $p_i = r_i + s_i$ and differences $u_i = r_i - s_i$ of the velocities for the Fotobalk and Counter instruments and compute $S_p^2 = 7.508$, $S_u^2 = 0.0590$, $S_{pu} = 0.1748$, so that from Eq. 2-61 $r_{pu} = 0.2626$, and from Eq. 2-62 we find

$$t(n-2,\ \sigma_{e_1} = \sigma_{e_2}) = r_{pu}\sqrt{n-2}/[1 - r_{pu}^2]^{1/2} = 0.861^{**}$$

for Student's t to compare σ_{e_2} and σ_{e_1}, whereas $t_{0.90}(10) = 1.372$ and $t_{0.95}(10) = 1.812$. We therefore conclude that the Fotobalk and Counter chronographs have equal precision of measurement, even though for 12 rounds $\hat{\sigma}_{e_1} = 0.081$ m/s and $\hat{\sigma}_{e_2} = 0.229$ m/s as indicated in Table 2-4. Had we used a much larger sample size, we possibly could have established that I_1 is much more precise than I_2 although we were not able to detect any difference in precision of measurement for the two instruments for only $n = 12$ observations.

2-6.3 TEST OF WHETHER $\beta_1 = \beta_2$ (ACCURACY TEST)

Next we check the agreement in the true unknown levels of measurement for the Fotobalk and Counter. This step is clearly and easily accomplished by using the differences in readings of I_1 and I_2, or $u_i = r_i - s_i$ and computing Student's t from

$$t_0\,(n-1,\ \beta_1 = \beta_2) = \bar{u}\sqrt{n}/S_u \qquad (2\text{-}63)$$

$$= -0.608\sqrt{12}/(0.2429) = -8.67$$

*Velocity firings generally destroy the projectiles.

**The t value of 0.861 for 10 df actually corresponds to a probability of about 0.79.

TABLE 2-4

ESTIMATES OF PRECISION OF MEASUREMENT ON THREE SIMULTANEOUS VELOCITY MEASUREMENTS OF THE FOTOBALK, COUNTER, AND TERMA CHRONOGRAPHS (Ref. 12)

Round No.	Foto I_1 r	Counter I_2 s	Terma I_3 t	Mean Velocity, m/s	$r-s$ $=u$	$s-t$ $=v$	$t-r$ $=w$
20	793.8	794.6	793.2	793.87	−0.8	+1.4	−0.6
21	793.1	793.9	793.3	793.43	−0.8	+0.6	+0.2
22	792.4	793.2	792.6	792.73	−0.8	+0.6	+0.2
23	794.0	794.0	793.8	793.93	0.0	+0.2	−0.2
24	791.4	792.2	791.6	791.73	−0.8	+0.6	+0.2
25	792.4	793.1	791.6	792.37	−0.7	+1.5	−0.8
26	791.7	792.4	791.6	791.90	−0.7	+0.8	−0.1
27	792.3	792.8	792.4	792.50	−0.5	+0.4	+0.1
28	789.6	790.2	788.5	789.43	−0.6	+1.7	−1.1
29	794.4	795.0	794.7	794.70	−0.6	+0.3	+0.3
30	790.9	791.6	791.3	791.27	−0.7	+0.3	+0.4
31	793.5	793.8	793.5	793.60	−0.3	+0.3	0.0

$$S_u^2 = S_{r-s}^2 = 0.0590 \qquad \bar{u} = \beta_1 - \beta_2 + \bar{e}_1 - \bar{e}_2 = -0.608$$

$$S_v^2 = S_{s-t}^2 = 0.2711 \qquad \bar{v} = \beta_2 - \beta_3 + \bar{e}_2 - \bar{e}_3 = +0.725$$

$$S_w^2 = S_{t-r}^2 = 0.2252 \qquad \bar{w} = \beta_3 - \beta_1 + \bar{e}_3 - \bar{e}_1 = +0.117$$

$$\text{est}\sigma_{e_1}^2 = 0.5\,(0.0590 + 0.2252 - 0.2711) = 0.0065 \quad \text{(Eq. 2-49)}$$
$$\text{est}\sigma_{e_1} = 0.081 \text{ m/s}$$
(Foto)

$$\text{est}\sigma_{e_2}^2 = 0.5\,(0.0590 - 0.2252 + 0.2711) = 0.0525 \quad \text{(Eq. 2-50)}$$
$$\text{est}\sigma_{e_2} = 0.229 \text{ m/s}$$
(Counter)

$$\text{est}\sigma_{e_3}^2 = 0.5\,(-0.0590 + 0.2252 + 0.2711) = 0.2186 \quad \text{(Eq. 2-51)}$$
$$\text{est}\sigma_{e_3} = 0.468 \text{ m/s}$$
(Terma)

$\text{est}\sigma_x = 1.42$ m/s = estimated standard deviation of the true velocities of the rounds (Eq. 2-56).

Reprinted with permission. Copyright © by American Statistical Association.

which for $(n-1) = 11$ df is very highly significant (t_o is the observed value of t). Thus we would look for the cause of this disagreement, i.e., run a retest of the two "standards" or calibrate them since the Fotobalk reads 0.61 m/s lower than the Counter. In this case, however, the sample variance of the differences in errors of measurement is very small, i.e., $S_u^2 = 0.0590$ (or $S_u = 0.2429$), and our t test is sensitive enough to pick up easily a difference of 0.61 m/s in velocity levels. It could happen, for example, that the Fotobalk might be found, through more testing, to be more precise than the Counter and hence could be easier to calibrate. Also, in the absence of any further data, we might recognize and correct for the apparent difference of 0.61 m/s. (The correct direction is unknown!)

2-6.4 LARGE SAMPLE TEST OF WHETHER σ_{e_1} or σ_{e_2} EQUALS ZERO

Maloney and Rastogi (Ref. 10) point out that for large sample size n Wilks' (Ref. 13) likelihood ratio test may be used to detect whether σ_{e_1} or σ_{e_2} can be considered to be zero. To test the hypothesis that $\sigma_{e_1} = 0$, for example, they point out that the likelihood ratio λ is

$$\lambda = \{(S_r^2 S_s^2 - S_{rs}^2) \,/\, [S_r^2(S_r^2 + S_s^2 - 2S_{rs})]\}^{n/2} \tag{2-64}$$

and according to Wilks (Ref. 13), then

$$-2\ln\lambda = \chi^2\,(1). \tag{2-65}$$

That is, $-2\ln\lambda$ follows the chi-square distribution with 1 df. If we desire to test whether $\sigma_{e_2} = 0$, the single factor S_r^2 before the brackets in the denominator of Eq. 2-64 would be replaced by S_s^2.

Example 2-5:

Return to the data of Table 2-2, where for the two-instrument case it seemed necessary to take $\sigma_{e_2} = 0$. Is there any evidence from the Maloney-Rastogi test to conclude that actually $\sigma_{e_2} = 0$?

To answer this question, we have $n = 29$, $S_r^2 = 0.0467544$, $S_s^2 = 0.0451123$, and $S_{rs} = 0.0455819$. (We omitted the 10.01 of I_1 for which I_2 lost a round.) Hence from Eqs. 2-64 and 2-65

$$-2\ln\lambda = -2\ln[(0.000031489)/(0.000031709)]^{14.5} = 0.2019.$$

The observed value of $-2\ln\lambda = 0.2019$. Referring this value to a table of probability levels of $\chi^2(1)$, we find $P \approx 0.35$. Thus we must accept the null hypothesis that $\sigma_{e_2} = 0$ and conclude this is possible. We could not reject the null hypothesis that $\sigma_{e_2} = 0$ unless the value of S_s^2, substituted for the single S_r^2 in Eq. 2-64, would give a value calculated by Eq. 2-64 exceeding the upper 5% level of $\chi^2(1)$.

2-6.5 TEST FOR WHETHER $\sigma_{e_2} = k\sigma_{e_1}$ AND SHUKLA's TEST

We return to the significance test of Eq. 2-62 for the two-instrument case where we test whether $\sigma_{e_1} = \sigma_{e_2}$ or whether the true population correlation coefficient of Eq. 2-60 is $\rho = 0$. Our procedure is actually to assume $\rho = 0$; to calculate the observed or sample correlation coefficient r_{pu} in Eq. 2-61; and then refer this value to a table of the null distribution of r_{pu}, or use Eq. 2-62, to determine whether it is significant. Similarly, we may assume or hypothesize any value of ρ for $-1 < \rho < 1$, $\rho \neq 0$; calculate the sample r_{pu}; and then refer the latter calculated value to the proper table of $r = r_{pu}$ for the assumed value of $\rho \neq 0$. This means that the hypothesized value of ρ is calculated from Eq. 2-60 with, for example, σ_{e_1} and σ_x as specified multiples of σ_{e_2}, etc.

An alternative, approximate procedure is to calculate

$$\frac{\sqrt{n-3}}{2}\{\ln[(1+r)/(1-r)] - \ln[(1+\rho)/(1-\rho)]\} \approx N(0,1) \tag{2-66}$$

which for large sample size n has been shown by R. A. Fisher to be approximately normally distributed with zero mean and unit standard deviation.

We may obtain a "numerical calibration" of the value of Eq. 2-66 for small n by making a calculation relative to Example 2-4 and the data of Table 2-4 for I_1 and I_2. We found that the observed $r_{pu} = 0.2626$, and for $n = 12$ with the assumption $\rho = 0$, the left-hand side (LHS) of Eq. 2-66 is

$$\frac{\sqrt{12-3}}{2}[\ln(1 + 0.2626)/(1 - 0.2626)] = 0.81$$

which, when referred to a normal probability table, gives a chance of 0.79* for a one-sided test or 0.58 for the two-sided test, a very accurate value for $n = 12$! By examining Eq. 2-60, it is seen that if the product variability $\sigma_x = 0$, the population correlation coefficient ρ becomes

$$\rho_{pu} = (\sigma_{e_1}^2 - \sigma_{e_2}^2)/(\sigma_{e_1}^2 + \sigma_{e_2}^2). \tag{2-67}$$

In this case the significance test based on the observed sample correlation coefficient $r = r_{pu}$ would be very sensitive to unequal (or equal) σ_{e_1} and σ_{e_2}. Otherwise, as σ_x approaches larger and larger values relative to σ_{e_1} and σ_{e_2}, the product variability dominates Eqs. 2-60 and 2-61, so that the ratio $\sigma_{e_2}/\sigma_{e_1} = k$ becomes obscured and the test becomes insensitive.

Example 2-6:

With reference to Example 2-4 and the data of Table 2-4, is it reasonable to conclude that we could have a highly distorted ratio such as $\sigma_{e_2} = 9\sigma_{e_1}$ when we take the product variability to be $\sigma_x = 1.42$ m/s and hence show test insensitivity?

We could estimate that $\sigma_x/\sigma_{e_2} \approx 1.42/0.229 = 6.20$ or $\sigma_x = 55.8\sigma_{e_1}$, which is large indeed, and substituting this value and the assumption $\sigma_{e_2} = 9\sigma_{e_1}$ into Eq. 2-60, we calculate est$\rho = \rho_{pu} \approx -0.0789$, a near zero value.

The sample correlation coefficient in Example 2-4 was calculated to be

$$r = r_{pu} = 0.2626.$$

Hence from Eq. 2-66

$$\frac{\sqrt{12-3}}{2}\left[\ln\left(\frac{1+0.2626}{1-0.2626}\right) - \ln\left(\frac{1-0.0789}{1+0.0789}\right)\right] = 1.04$$

which, when referred to a table of the standardized normal probability integral, gives an insignificant probability P of $P \approx 0.85$ (one-sided). Consequently, we do not reject the null hypothesis that perhaps the ratio $\sigma_{e_2} = 9\sigma_{e_1}$ could be true!

Shukla (Ref. 14) has proposed a very clever test concerning whether $\sigma_{e_2}^2 = k^2\sigma_{e_1}^2$ and has thus generalized the Maloney-Rastogi (Ref. 10) test. Shukla (Ref. 14) puts

$$u_i = r_i - s_i, \text{ as we do in Eq. 2-47,}$$

but

$$g_i = s_i + k^2 r_i \tag{2-68}$$

where we call our instrument I_2 Shukla's 1. For this formulation the population correlation coefficient ρ of Eq. 2-60 is changed to

$$\rho = \frac{\sigma_{e_2}^2 - k^2\sigma_{e_1}^2}{\{(\sigma_{e_1}^2 + \sigma_{e_2}^2)\,[\sigma_{e_2}^2 + k^4\sigma_{e_1}^2 + \sigma_x^2\,(1+k^2)^2]\}^{1/2}} \tag{2-69}$$

and the observed sample correlation coefficient $r = r_{ug}$** between the random variables u and g in terms of the original instrument readings r_i and s_i is

$$r = r_{ug} = \frac{S_{ug}}{\sqrt{S_u^2 S_g^2}} = \frac{S_s^2 - k^2 S_r^2 + (k^2-1)S_{rs}}{[(S_r^2 + S_s^2 - 2S_{rs})\,(S_s^2 + k^4 S_r^2 + 2k^2 S_{rs})]^{1/2}}. \tag{2-70}$$

*For a two-sided test, the chance would be 0.58, which would usually be more appropriate.

**Although r_i is used in this chapter as an instrumental reading, the notation "r" is widely used as a correlation coefficient.

Hence to test the null hypothesis that $\sigma_{e_2} = k\sigma_{e_1}$, we also use the Student's t test as did Maloney and Rastogi (Ref. 10), or the first form on which our Eq. 2-62 is based, i.e.,

$$t(n-2,\ \sigma_{e_2} = k\sigma_{e_1}) = \frac{r_{ug}\sqrt{n-2}}{(1-r_{ug}^2)^{1/2}} . \tag{2-71}$$

When $k = 1$, the Shukla test (Ref. 14) is precisely that of Maloney and Rastogi (Ref. 10). Putting $k = 0$ tests whether $\sigma_{e_2} = 0$. (We note also that when $k = 1$, Eq. 2-69 becomes the negative of Eq. 2-60. This change in sign is due to our switching instruments in Shukla's notation to test our $\sigma_{e_2} = k\sigma_{e_1}$.)

We will use Shukla's test to judge whether $\sigma_{e_2} = 9\sigma_{e_1}$, or, that is, solve Example 2 a different way. We calculate

$$S_r^2 = 1.9790, \qquad S_s^2 = 1.8042, \qquad S_{rs} = 1.8621.$$

Then with $k = 9$ we find from Eq. 2-70 that

$$r = -0.340$$

and from Eq. 2-71,

$$t = -1.14$$

which is not significant at the 0.05 level since $t_{0.05} = -1.812$. Thus we cannot reject the stated hypothesis $\sigma_{e_2} = 9\sigma_{e_1}$ with Shukla's test either! (This again demonstrates test insensitivity!)

So far for the two-instrument case, we have accepted the null hypothesis that $\sigma_{e_1} = \sigma_{e_2}$ and that $\sigma_{e_1} \neq 0$; now we have also accepted the hypothesis that $\sigma_{e_2} = 9\sigma_{e_1}$! This certainly amounts to some unpleasant contradictions, but perhaps it also possibly indicates the relative insensitivity of significance tests to the components of variance studied here, especially for small n and estσ_{e_1} near zero. More will be said about this problem for the three-instrument case, for which we will demonstrate also that perhaps much larger sample sizes may be required.

2-7 SIGNIFICANCE TESTS FOR THREE INSTRUMENTS*

2-7.1 INTRODUCTORY REMARKS

Having seen some problems with estimation and significance tests of precision and accuracy for only two instruments, especially since the product variability might mask desired comparisons, we now examine some appropriate statistical tests of hypotheses for measurements with three instruments—I_1, I_2, and I_3. For the three-instrument case we saw that the estimation of precision and accuracy parameters turned out to be very favorable indeed and no doubt worthwhile.

For the three-instrument case several statistical tests of significance are available that appear to be very useful indeed. We should, however, pause to reflect on just which statistical tests would be the more desirable ones. In view of the masking problem caused by product variation for two instruments, it certainly seems desirable to use three instruments for determining whether $\sigma_{e_1} = \sigma_{e_2}$ for the first two "designated" instruments without regard to the imprecision σ_{e_3} for the third instrument. Also there is the problem of being able to determine just which of the three instruments is the "best" or the "worst", so to speak. Therefore, it becomes desirable to make comparisons of one instrument versus the other two. This leads to using or establishing two of the instruments as "reference" or "standard" instruments to test the "worth" of the third instrument. In fact, this may become especially desirable whenever we are dealing with small sample sizes or until we can actually obtain enough valid information on precision and accuracy to depend on two of the instruments as good reference or standard ones. Finally, there will be some need occasionally to test composite hypotheses concerning all three instruments and their precision and accuracy capabilities. We will start with a test of whether $\sigma_{e_1} = \sigma_{e_2}$ using data for all three instruments.

* For a recent development in testing the equality of three instrumental imprecisions, please see par. 2-12, "Additional Discussion".

2-7.2 THREE-INSTRUMENT TEST OF WHETHER $\sigma_{e_1} = \sigma_{e_2}$

For this case and the assumption of normally distributed uncorrelated errors of measurement, Grubbs (Ref. 12) has shown that the appropriate test based on Student's t is

$$t(n-2,\ \sigma_{e_1} = \sigma_{e_2}) = \frac{[(S_v^2/S_w^2) - \theta]\ (n-2)^{1/2}}{[4\theta(1 - r_{vw}^2)\ (S_v^2/S_w^2)]^{1/2}}. \tag{2-72}$$

where

θ = ratio of the expected values of the variances of $v = s - t$ and $w = t - r$

and hence is clearly

$$\theta = (\sigma_{e_2}^2 + \sigma_{e_3}^2)/(\sigma_{e_1}^2 + \sigma_{e_3}^2). \tag{2-73}$$

Hence a test of whether $\sigma_{e_1} = \sigma_{e_2}$ or whether I_1 and I_2 are equally precise is also the test of whether $\theta = 1$ in Eq. 2-73.

Example 2-7:

Referring to Example 2-4 and the data of Table 2-4, where only the Fotobalk and the Counter were used to determine whether $\sigma_{e_1} = \sigma_{e_2}$, we now use available data for all three velocity chronographs (including the Terma) to test whether $\sigma_{e_1} = \sigma_{e_2}$.

By substituting in Eq. 2-72 we calculate

$$t(n-2,\ \sigma_{e_1} = \sigma_{e_2}) = \frac{[(0.2711)/(0.2252) - 1]\ \sqrt{10}}{\{4[1 - (0.8847)^2]\ (0.2711)/(0.2252)\}^{1/2}}$$

$$= 0.63 \qquad (t_{0.95} = 1.812)$$

which is not a significant value of t for 10 df. We conclude, therefore, that for the more precise test of the three-instrument case and for $n = 12$ rounds, we do not reject the hypothesis that $\sigma_{e_1} = \sigma_{e_2}$ or that the Fotobalk and Counter possess equivalent precision of measurement. This result seems to substantiate the need for a larger sample size.

2-7.3 THREE-INSTRUMENT TEST OF WHETHER $\sigma_{e_2} = k\sigma_{e_1}$ (SHUKLA'S TEST)

Shukla (Ref. 15) has developed an apparently powerful test of whether $\sigma_{e_2} = k\sigma_{e_1}$ when three instruments are used. This Shukla test (Ref. 15) uses

$$u_i = r_i - s_i, \text{ as in Eq. 2-47}$$

$$v_i = s_i - t_i, \text{ as in Eq. 2-47}$$

and then takes

$$h_i = u_i + (\delta + 1)\ v_i \tag{2-74}$$

where

$$\delta = 1/k^2.$$

Then the sample correlation coefficient between the random variables u and h is for $\delta = 1/k^2$

$$r = r_{uh} = S_{uh}/(S_u^2 S_h^2)^{1/2} \tag{2-75}$$

$$= [S_u^2 + (\delta + 1)S_{uv}]\{S_u^2[S_u^2 + 2(\delta + 1)S_{uv} + (\delta + 1)^2 S_v^2]\}^{-1/2}. \tag{2-76}$$

Again this leads to use of the Student's t test, or

$$t(n - 2, \sigma_{e_2} = k\sigma_{e_1}) = \frac{r_{uh}\sqrt{n-2}}{(1 - r_{uh}^2)^{1/2}}. \tag{2-77}$$

Example 2-8:

In Example 2-6 we carried out some two-instrument tests of whether $\sigma_{e_2} = 9\sigma_{e_1}$ and concluded for $n = 12$ rounds that we could not reject this hypothesis nor could we reject $\sigma_{e_2} = \sigma_{e_1}$. In view of Shukla's more precise or powerful three-instrument test, apply it to determine whether we may conclude that $\sigma_{e_2} = 9\sigma_{e_1}$.

We have

$$S_u^2 = 0.05902, \qquad S_v^2 = 0.27114, \qquad S_{uv} = -0.0525$$

$$\delta = 1/9^2 = 1/81 = 0.01235$$

and from Eq. 2-76 we find

$$r = r_{uh} = 0.00068499$$

and the Student's t of Eq. 2-77 is

$$t(10, \sigma_{e_2} = 9\sigma_{e_1}) = 0.00217.$$

Again this is not a significant value of t, so we must conclude from the more sensitive Shukla's three-instrument test that we cannot reject the hypothesis that $\sigma_{e_2} = 9\sigma_{e_1}$!

The result of this test, using data for all three instruments, actually confirms our findings for the use of only two instruments. Accordingly, we probably should have more confidence or assurance that the two-instrument test of Shukla's in par. 2-6.5 is really not too insensitive for departures from the assumptions or hypothesized values about the ratio of large sample or population imprecisions σ_{e_1} and σ_{e_2}.

In summary, for both the two- and three-instrument cases, we have insufficient information to reject that $\sigma_{e_1} = \sigma_{e_2}$—i.e., that I_1 and I_2 are equally precise—and moreover, we have insufficient evidence to reject that possibly $\sigma_{e_2} = 9\sigma_{e_1}$! Thus such questions probably could be settled by increasing the sample size or perhaps by use of a much more precise third instrument than the Terma. For example, better precision might result in the test of Eq. 2-72 if σ_{e_3} in Eq. 2-73 were much smaller or even for the Shukla test of Eq. 2-77 if we had a very precise third or standard instrument. Finally, the reader may appreciate that we have selected an example that shows some possible difficulties one should expect for certain precision and accuracy tests along with the probable requirement to perform sufficiently extensive calibration.

We have some reservations about the Fotobalk and Counter being compatible as reference or standard instruments because we found a significant difference in instrumental biases, and there also is some sample estimation evidence that perhaps σ_{e_2} may be as large as about $9\sigma_{e_1}$; this perhaps is obscured by σ_{e_3}. We should, though, continue to accumulate precision and accuracy data. However, this need not be a concern in what follows, for as it turns out we may compare the precision of measurement of the Terma with the average precision of the Fotobalk and Counter and the bias of the Terma with the average bias of the Fotobalk and Counter—a desirable procedure.

2-7.4 JUDGMENT PROCEDURES FOR TESTING A THIRD INSTRUMENT

We will proceed to indicate the applicable significance test procedures to determine whether or not a third or "test" instrument should be "accepted". In particular, we will consider the Fotobalk and the Counter as standard or reference instruments—until we get better ones or have more experience—and will proceed to determine the usefulness of the Terma chronograph. The suitability of the Terma instrument will be assessed by studying whether it is as precise and as accurate as the Fotobalk and Counter chronographs. The procedures discussed are covered thoroughly in Ref. 12, and the reader should examine the computations in Table 2-5, where the sums (less a convenient origin, such as 1580) and differences of the two reference instrument observations are given along with the differences in readings between the Terma or "test" instrument and the average of the two standard instrument readings. Also certain correlation coefficients are calculated for use as described in the significance tests that follow on precision and accuracy of the Terma versus the "average" of the Fotobalk and Counter.

To ascertain whether the variance in errors of measurement of the Terma chronograph is equal to that of the average of the Fotobalk and Counter instruments, we use Ref. 12 and put

$$\nu = [\sigma_{e_3}^2 + (\sigma_{e_1}^2 + \sigma_{e_2}^2)/4]/(\sigma_{e_1}^2 + \sigma_{e_2}^2) = 3/4$$

in the statistic

$$t_0[n-2,\ \sigma_{e_3}^2 = (\sigma_{e_1}^2 + \sigma_{e_2}^2)/2] = \frac{(S_z^2/S_u^2 - \nu)\sqrt{n-2}}{[4\nu(1-r_{zu}^2)\,S_z^2/S_u^2]^{1/2}} \quad * \tag{2-78}$$

$$= \frac{[(0.2334)/(0.0590) - 0.75]\sqrt{10}}{\{3[1-(0.1959)^2]\,(0.2334)/(0.0590)\}^{1/2}}$$

$$= 3.00.$$

We therefore conclude that the Terma chronograph is not as precise as the ("average" of the) Fotobalk and Counter instruments since $t_{0.95}\,(10) = 1.812$.

We note from Table 2-4 that the standard deviation in errors of measurement for the Terma chronograph is estimated as 0.468 m/s, and this instrument is measuring an estimated standard deviation in true velocity of 1.42 m/s, so that it is of questionable precision for the measurements taken here. Nevertheless, we may want to check on the speed measured by the Terma chronograph, which may be determined by using the lst column of Table 2-5 and calculating

$$t[n-1,\ \beta_3 = (\beta_1 + \beta_2)/2] = \bar{z}\sqrt{n}/S_z \tag{2-79}$$

$$= -0.421\sqrt{12}/0.483 = -3.02.$$

Since $t_{0.95}(11) = 1.796$, we conclude that the Terma chronograph reads low by 0.421 m/s as compared to the average of the Fotobalk and Counter. (Note that the bias of 0.61 m/s between the two "standards" is even a bit larger.)

The variance in errors of measurement of the Terma or third chronograph may be estimated also from

$$\text{est}\sigma_{e_3}^2 = S_z^2 - S_u^2/4 \tag{2-80}$$

$$= 0.2334 - 0.0590/4 = 0.2187 \text{ m/s},$$

*See Eq. 2-92 for the general value of ν.

TABLE 2-5

SIMULTANEOUS VELOCITIES OF THE FOTOBALK, COUNTER, AND TERMA CHRONOGRAPHS WITH TEST VS STANDARD COMPARATIVE DATA ON EACH OF TWELVE SUCCESSIVE ROUNDS, m/s (Ref. 12)

Round No.	Foto r I_1	Counter s I_2	Terma t I_3	$(r+s)-1580=y$	$r-s=u$	$t-(r+s)/2=z$
20	793.8	794.6	793.2	8.4	−0.8	−1.00
21	793.1	793.9	793.3	7.0	−0.8	−0.20
22	792.4	793.2	792.6	5.6	−0.8	−0.20
23	794.0	794.0	793.8	8.0	0.0	−0.20
24	791.4	792.2	791.6	3.6	−0.8	−0.20
25	792.4	793.1	791.6	5.5	−0.7	−1.15
26	791.7	792.4	791.6	4.1	−0.7	−0.45
27	792.3	792.8	792.4	5.1	−0.5	−0.15
28	789.6	790.2	788.5	−0.2	−0.6	−1.40
29	794.4	795.0	794.7	9.4	−0.6	0.00
30	790.9	791.6	791.3	2.5	−0.7	+0.05
31	793.5	793.8	793.5	7.3	−0.3	−0.15

$$S_y^2 = [n\Sigma y_i^2 - (\Sigma y_i)^2]/[n(n-1)] = [12(448.89) - (66.3)^2]/132 = 7.508$$

$$S_u^2 = [12(5.09) - (-7.3)^2]/132 = 0.0590$$

$$S_z^2 = [12(4.6925) - (-5.05)^2]/132 = 0.2334$$

$$S(z) = 0.483$$

$$S_{yu} = [n\Sigma y_i u_i - (\Sigma y_i)(\Sigma u_i)]/[n(n-1)] = [12(-38.41) - (66.3)(-7.3)]/132 = 0.1748$$

$$S_{uz} = [12(3.325) - (5.05)(7.3)]/132 = 0.0230$$

$$r_{yu} \doteq S_{yu}/(S_u^2 S_y^2)^{1/2}$$

$$r_{yu} = (0.1748)/\sqrt{(7.508)(0.0590)} = 0.2626$$

$$r_{uz} = (0.0230)/\sqrt{(0.2334)(0.0590)} = 0.1959$$

Mean $(r-s) = \bar{u} = -0.608$ m/s

$\bar{z} = -0.421$ m/s

Reprinted with permission. Copyright © by American Statistical Association.

which agrees with the value of 0.2186 computed by the equivalent equation in Table 2-4. Hence estσ_{e_3} = 0.468.

The standard deviation of the mean velocities listed in the fifth column of Table 2-4—i.e., from the model $x_i + (\beta_1 + \beta_2 + \beta_3)/3 + (e_{i1} + e_{i2} + e_{i3})/3$—is found to be 1.43 m/s as compared to the estimated true value of 1.42 m/s. Therefore, we conclude that the variance in errors of each measuring instrument is appreciably smaller than the (population) variance of the velocities of the rounds. Nevertheless, some

calibration of instruments may be highly desirable or even required. In addition, appropriate information should continue to be acquired to designate finally standard or reference (very dependable) instruments for calibrations and other purposes.

The theory of Ref. 12 should be generalized to provide a significant ranking of any number of measuring instruments with regard to both precision and accuracy. This would amount to a very important and highly practical accomplishment indeed. It is highly desirable that the significance tests developed should point out the particular instruments that are relatively imprecise or inaccurate, as was attempted here.

2-8 CONFIDENCE BOUNDS ON THE UNKNOWN PRECISION AND ACCURACY PARAMETERS, AND ALLIED ACCOMPLISHMENTS

Since we have developed several appropriate statistical significance tests concerning the unknown precision and accuracy parameters for two and three instruments, it becomes readily apparent to the reader that confidence bounds on certain of the parameters or functions of them may be easily established although the establishment of some others may be rather difficult.

2-8.1 CONFIDENCE BOUNDS ON $(\beta_1 - \beta_2)$ FOR TWO INSTRUMENTS

To begin with, it is easy to establish confidence bounds on the differences in biases between the pairs of instruments. In fact, for instruments I_1 and I_2 and the assumptions of normality and independence, we have that $\bar{u} = \bar{r} - \bar{s}$ is normally distributed with mean $(\beta_1 - \beta_2)$ and variance equal to $(\sigma_{e_1}^2 + \sigma_{e_2}^2)/n$, which involves the imprecisions and sample size n. Thus using Student's t distribution with $(n - 1)$ df or Eq. 2-63, the $(1 - 2\alpha)$ confidence bounds on the true unknown difference $(\beta_1 - \beta_2)$ in biases of I_1 and I_2 are found from

$$Pr\,[\bar{u} - \sqrt{n}\, t_{1-\alpha}/S_u \leq \beta_1 - \beta_2 \leq \bar{u} + \sqrt{n}\, t_{1-\alpha}/S_u] = 1 - 2\alpha. \tag{2-81}$$

Also either a lower or an upper one-sided $(1 - \alpha)$ confidence bound on $(\beta_1 - \beta_2)$ is clearly obtainable from the end points of Eq. 2-81.

2-8.2 CONFIDENCE BOUNDS ON $[\beta_3 - (\beta_1 + \beta_2)/2]$ FOR THREE INSTRUMENTS

In a manner very similar to that of par. 2-8.1, it can be seen—using $z = t - (r + s)/2$, i.e., the last column of Table 2-5—that the $(1 - 2\alpha)$ confidence bounds on the difference $[\beta_3 - (\beta_1 + \beta_2)/2]$ between the bias of the third instrument and the average bias of the first two instruments are found from

$$Pr\,[\bar{z} - \sqrt{n}\, t_{1-\alpha}/S_z \leq \beta_3 - (\beta_1 + \beta_2)/2 \leq \bar{z} + \sqrt{n}\, t_{1-\alpha}/S_z] = 1 - 2\alpha. \tag{2-82}$$

or alternatively an upper or a lower $(1 - \alpha)$ confidence bound. Student's t with $(n - 1)$ df is used.

2-8.3 PRELIMINARY COMMENTS ON CONFIDENCE BOUNDS FOR PRECISION PARAMETERS

Whereas confidence bounds are easily established on the true differences in instrumental biases or systematic errors, the theory is more complicated for the unknown precision parameters. To begin with, the functional forms of the precision parameters are much more complex, and some nuisance parameters are present, which make the problem analytically troublesome. In some cases, therefore, some calculations may be carried out only when absolutely necessary or perhaps as a last resort. However, we will at least indicate some of the problems involved and show how confidence bounds may be obtained for several important cases. These statements apply primarily to confidence bounds on the desired ratios, such as $\sigma_{e_2}/\sigma_{e_1}$. Fortunately, as a result of rather intensive research in recent years, simultaneous confidence bounds or regions for all of the parameters jointly can be found by the methods of multivariate statistical analysis. We will give a brief account of useful results and will refer to the appropriate literature on the subject.

2-8.4 CONFIDENCE BOUNDS ON PRECISION PARAMETERS FOR TWO INSTRUMENTS

A lower $(1 - \alpha)$ confidence bound on the relative precision of measurement, or ratio σ_x/σ_{e_1}, is readily available from Eq. 2-37. An upper $(1 - \alpha)$ bound is found by changing signs of the $t_{1-\alpha}$'s in Eq. 2-37. This upper bound is taken as infinity if the denominator is negative or zero. (The same is true for σ_x/σ_{e_2}.)

Confidence bounds on the population correlation coefficient of Eq. 2-60 may be found by using an appropriate Student's t statistic or even the normal approximation of Eq. 2-66. However, we note in Eq. 2-60 that there is the nuisance parameter σ_x and that confidence bounds on the desired ratio, say $\sigma_{e_2}/\sigma_{e_1}$, must be found by the Shukla method that follows. If I_1 and I_2 measure the same item n times, thereby making $\sigma_x = 0$, then suitable confidence bounds for the ratio of imprecisions could be established through the use of Eq. 2-67. In comparing only measuring instruments such a procedure may often be desired or even necessary as a simple, practical approach to studying precision of measurement (instrument capability).

For joint or simultaneous confidence bounds or regions on all parameters for the two-instrument case, including product variation, the results of Thompson (Refs. 3 and 16) are especially important and noteworthy. Indeed, using multivariate statistical methods Thompson shows, for the two-instrument case, that the probability is at least $(1 - 2\alpha)$ that the following three relations hold simultaneously:

$$|\sigma_x^2 - (n-1)S_{rs}K| \leq M(n-1)\,(S_r^2\,S_s^2)^{1/2} \tag{2-83}$$

$$|\sigma_{e_1}^2 - (n-1)\,(S_r^2 - S_{rs})\,K| \leq M(n-1)\,[S_r^2\,(S_r^2 + S_s^2 - 2S_{rs})]^{1/2} \tag{2-84}$$

$$|\sigma_{e_2}^2 - (n-1)\,(S_s^2 - S_{rs})\,K| \leq M(n-1)\,[S_s^2\,(S_r^2 + S_s^2 - 2S_{rs})]^{1/2} \tag{2-85}$$

where the factors K and M are found in Table 2-6 (Table 2 of Ref. 3) for $2\alpha = 0.01$ and $2\alpha = 0.05$.

Example 2-9:

Return to the data of Table 2-2 for the fuze burning times, and use all 30 readings of the first and third instruments (I_1 and I_3) to obtain simultaneous 95% confidence bounds on the standard deviations of product variability and the two imprecisions of measurement, i.e., σ_x, σ_{e_1}, and σ_{e_3}.

We calculate

$$S_r^2 = 0.04714 \qquad S_t^2 = 0.04561 \qquad S_{rt} = 0\ .04593$$

and note that $\text{est}\sigma_x = \sqrt{0.04593} = 0.214$, $\text{est}\sigma_{e_1} = \sqrt{0.04714 - 0.04593} = 0.0347$, but $\text{est}\sigma_{e_3} < 0$, and hence we must take $\text{est}\sigma_{e_3} = 0$ here also.

By substituting the calculated variances, the covariance, and the K and M of Table 2-6 for $2\alpha = 0.05$ into Eqs. 2-82, 2-83, and 2-84, we obtain with 95% confidence that simultaneously

$$\begin{aligned} 0.16 &< \sigma_x &&< 0.32 \\ 0.00 &< \sigma_{e_1} &&< 0.09 \\ 0.00 &< \sigma_{e_2} &&< 0.07. \end{aligned}$$

(All negative lower bounds must be replaced by zero.)

Finally, for the two-instrument case, confidence bounds on the ratio $\sigma_{e_2}/\sigma_{e_1}$ are obtainable as a result of the work by Shukla (Ref. 14). In fact, as shown by Shukla (Ref. 14), confidence bounds on the unknown ratio $\sigma_{e_2}/\sigma_{e_1} = k$ of population imprecisions may be found with the aid of Eqs. 2-70 and 2-71. Thus from Eq. 2-71 and for given upper and lower α probability levels for Student's t, corresponding bounds for r_{ug} may be determined. Then by using Eq. 2-70, the solution of a quadratic equation will give $(1 - 2\alpha)$ confidence bounds for k^2, from which the confidence bounds for $k = \sigma_{e_2}/\sigma_{e_1}$ may be obtained by taking square roots, as indicated by Eqs. 2-90 and 2-91.

TABLE 2-6

VALUES OF K AND M WHICH YIELD
$(1 - 2\alpha)$ CONFIDENCE REGIONS WHEN USED IN
CONJUNCTION WITH EQUATIONS 2-83 THROUGH 2-85 (Ref. 3)

$n-1$	$2\alpha = 0.01$		$2\alpha = 0.05$	
	K	M	K	M
3	99.78	99.72	19.79	19.71
4	12.38	12.33	4.146	4.077
5	3.980	3.931	1.726	1.665
6	1.903	1.858	0.9636	0.9083
7	1.120	1.078	0.6290	0.5786
8	0.7459	0.7076	0.4516	0.4052
9	0.5389	0.5031	0.3453	0.3022
10	0.4120	0.3782	0.2761	0.2357
11	0.3282	0.2963	0.2280	0.1901
12	0.2698	0.2395	0.1932	0.1573
13	0.2272	0.1983	0.1668	0.1328
14	0.1951	0.1675	0.1464	0.1140
15	0.1702	0.1438	0.1301	0.09925
16	0.1505	0.1251	0.1169	0.08738
17	0.1344	0.1100	0.1060	0.07767
18	0.1213	0.09772	0.09682	0.06962
19	0.1103	0.08752	0.08904	0.06287
20	0.1009	0.07896	0.08237	0.05713
22	0.08610	0.06546	0.07152	0.04795
24	0.07484	0.05538	0.06311	0.04098
26	0.06605	0.04763	0.05641	0.03554
28	0.05901	0.04152	0.05096	0.03121
30	0.05328	0.03660	0.04644	0.02768
35	0.04272	0.02778	0.03796	0.02127
40	0.03556	0.02200	0.03205	0.01700
45	0.03040	0.01797	0.02771	0.01398
50	0.02652	0.01503	0.02440	0.01176
60	0.02109	0.01110	0.01967	0.00875
70	0.01748	0.00862	0.01646	0.00684
80	0.01492	0.00694	0.01415	0.00553
90	0.01300	0.00575	0.01241	0.00460
100	0.01152	0.00486	0.01104	0.00390

Reprinted with permission. Copyright © by American Statistical Association.

If we let

$$F = S_s^2 - S_{rs} \tag{2-86}$$

$$G = S_r^2 - S_{rs} \tag{2-87}$$

$$H = t_\alpha^2 (S_r^2 S_s^2 - S_{rs}^2)/(n-2) \tag{2-88}$$

where

t_α = upper α probability level of Student's t

then Shukla (Ref. 14) has shown that the $(1 - 2\alpha)$ confidence bounds on $\sigma_{e_2}/\sigma_{e_1}$ are

$$Pr[D_L \leq \sigma_{e_2}/\sigma_{e_1} \leq D_U] = 1 - 2\alpha \tag{2-89}$$

where

$$D_L = [(G - \sqrt{H})/(F + \sqrt{H})]^{1/2} \tag{2-90}$$

$$D_U = [(G + \sqrt{H}]/[F - \sqrt{H})]^{1/2}. \tag{2-91}$$

Due to the possible existence of a negative F or G, especially for small sample sizes, the lower bound may have to be taken as zero, and the upper bound considered not calculable unless $F > \sqrt{H}$.

2-8.5 CONFIDENCE BOUNDS ON PRECISION PARAMETERS FOR THREE INSTRUMENTS

2-8.5.1 Confidence Bounds on $\sigma_{e_3}^2/[(\sigma_{e_1}^2 + \sigma_{e_2}^2)/2]$

When dealing with the data from three instruments, we can expect to obtain somewhat narrower confidence bounds on the unknown precision parameters than we can for only two instruments. In addition, it seems highly desirable in practice to compare one of the instruments to the other two. In fact, it will be most desirable, or even sometimes mandatory, to have access to at least two reference or standard instruments. We may then compare, as in par. 2-7.4, or place confidence bounds on the ratio of the precision of measurement of the "test" instrument with the average of the other two (reference) instruments. By referring to Eq. 2-78, for which we may select an upper and/or lower probability level for t, and with the sample data substituted therein, it can be seen that we may solve a quadratic equation in terms of the unknown parameter $\sqrt{\nu}$ from which upper and/or lower confidence bounds on ν are determined. Finally, since

$$\nu = \sigma_z^2/\sigma_u^2 = [\sigma_{e_3}^2 + (\sigma_{e_1}^2 + \sigma_{e_2}^2)/4]/\ (\sigma_{e_1}^2 + \sigma_{e_2}^2) \tag{2-92}$$

or

$$\sigma_{e_3}^2/[(\sigma_{e_1}^2 + \sigma_{e_2}^2)/2] = 2\nu - \tfrac{1}{2} \tag{2-93}$$

confidence bounds may be obtained for the LHS of Eq. 2-93, which is our goal. This will usually be done numerically as required on the part of the user.

2-8.5.2 Simultaneous Confidence Bounds On All Unknown Precision Parameters

Simultaneous confidence bounds on all of the precision parameters σ_{e_1}, σ_{e_2}, σ_{e_3}, and σ_x for the three-instrument case are available from multivariate statistical theory, as was the case in par. 2-8.4 for only two measuring instruments. In fact, the subject confidence bounds depend on percentage points (probability levels) of the extreme roots of a Wishart (multivariate) matrix as developed and calculated by Hanumara and Thompson (Ref. 17). As indicated by Hanumara and Thompson (Ref. 17), some of their work was stimulated by the original, practical problems of estimation of precision developed in Ref. 2. Fortunately, percentage points of the extreme roots of the pertinent Wishart matrix for cases involving 2, 3, 4, 5, 6, 7, 8, 9, and 10 instruments have been calculated by Hanumara and Thompson and are available in their Table 1 of Ref. 17. The sample sizes covered for three instruments are

$$n = 3(1)10(5)30(10)100$$

and the upper (u) and lower (ℓ) percentage points include probability levels of 0.005, 0.010, 0.025, and 0.050.

To indicate how computations of confidence bounds will be carried out, we need to express convenient multivariate notation. For any general number $N \geq 3$ of instruments, define the covariance of the n readings of any instruments j and k as

$$S_{jk} = \sum_{i=1}^{n} (r_{ij} - \bar{r}_{\cdot j})(r_{ik} - \bar{r}_{\cdot k})/(n-1) \tag{2-94}$$

where

r_{ij} = ith reading of instrument $j = 1, 2, \ldots, N$
r_{ik} = ith reading of instrument $k = 1, 2, \ldots, N$
$\bar{r}_{\cdot j}$ = sample mean of the readings of instrument j
$\bar{r}_{\cdot k}$ = sample mean of the readings of instrument k.

(We note with this notation that the sample variance of readings of the jth instrument is S_{jj}.) Using the preceding notation, Hanumara and Thompson (Ref. 17) show that for $N \geq 2$ instruments the probability is at least $(1 - 2\alpha)$ that the following confidence bounds obtain:

$$\frac{1}{2} \max_{j \neq k} [(n-1) S_{jk} (\ell^{-1} + u^{-1}) - (n-1)(\ell^{-1} - u^{-1})(S_{jj} S_{kk})^{1/2}]$$

$$\leq \sigma_x^2 \leq$$

$$\frac{1}{2} \min_{j \neq k} [(n-1) S_{jk} (\ell^{-1} + u^{-1}) + (n-1)(\ell^{-1} - u^{-1})(S_{jj} S_{kk})^{1/2}] \tag{2-95}$$

and

$$\frac{1}{2} \max_{j \neq 1} \{(n-1)(S_{11} - S_{1j})(\ell^{-1} - u^{-1}) - (n-1)(\ell^{-1} - u^{-1})[S_{11}(S_{11} + S_{jj} - 2S_{1j})]^{1/2}\}$$

$$\leq \sigma_{e_1}^2 \leq$$

$$\frac{1}{2} \min_{j \neq 1} \{(n-1)(S_{11} - S_{1j})(\ell^{-1} - u^{-1}) + (n-1)(\ell^{-1} - u^{-1})[S_{11}(S_{11} + S_{jj} - 2S_{1j})]^{1/2}\} \tag{2-96}$$

plus similar inequalities for $\sigma_{e_2}^2, \sigma_{e_3}^2, \ldots, \sigma_{e_N}^2$.

With $n = 12$, the data of Table 2-4 (or Table 2-5), and approximately interpolated lower (ℓ) and upper (u) $\alpha = 0.025\%$ points from Table 1 of Ref. 17, i.e.,

$$\ell \approx 2.01 \text{ and } u \approx 31.5,$$

the simultaneous 95% confidence bounds on the parameters are found to be

$0.77 \leq \sigma_x \leq 3.57$ m/s
$0.00 \leq \sigma_{e_1} \leq 0.92$
$0.00 \leq \sigma_{e_2} \leq 1.22$ m/s
$0.00 \leq \sigma_{e_3} \leq 1.98$.

Note how seemingly wide the 95% confidence bounds on the imprecisions of measurement appear to be for $n = 12$ rounds only.

2-8.5.3 Duplicate Measurements With One of Two Instruments and Allied Results

A very interesting and special case occurs if the readings or measurements of I_2, say, are replaced by duplicate determinations with instrument I_1. In other words, there are only two instruments really, with

one of them taking duplicate measurements. This is the case studied by Hahn and Nelson (Ref. 18), and it is readily seen that $\sigma_{e_1} = \sigma_{e_2} = \sigma_e$. Moreover, the quantity

$$S_z^2 / S_u^2 = F(n-1, n-1, \sigma_{e_3} = \sigma_e) \tag{2-97}$$

follows the Snedecor-Fisher F distribution with $(n-1)$ and $(n-1)$ df as indicated in Ref. 12. In addition, it is easy to establish that the lower and upper $(1-2\alpha)$ confidence bounds on $\sigma_{e_3}/\sigma_{e_1}$ are, respectively,

$$\{2S_z^2 / [(F_{1-\alpha}(n-1, n-1)S_u^2]\} - 1/2 \tag{2-98}$$

and

$$\{[2S_z^2 F_{1-\alpha}(n-1, n-1)] / S_u^2\} - 1/2. \tag{2-99}$$

We would especially recommend the continual acquisition of data on as many instruments as possible and the eventual accumulation of enough information to establish the precision parameters σ_{e_j} and the biases β_j or relative differences $(\beta_1 - \beta_2)$, etc., as accurately as possible. With such determination of stable estimates, one may make a valid selection of the more precise instruments for reference purposes or standards. In addition, there seems to be some advantage in selecting at least two instruments with small and equal imprecisions, e.g., $\sigma_{e_1} = \sigma_{e_2} = \sigma_e$, say. In such a situation, if we refer to the measurements I_1 and I_2 and consider their difference $u = r - s$ along with the quantity

$$z = -(s-t)/2 + (t-r)/2 = t - (r+s)/2 \tag{2-100}$$

then

$$4S_z^2/(3S_u^2) = F(n-1, n-1) \tag{2-101}$$

if $\sigma_{e_1} = \sigma_{e_2} = \sigma_e$. That is to say, the quantity $4S_z^2/(3S_u^2)$ follows the Snedecor-Fisher F distribution with $(n-1)$ and $(n-1)$ df. Hence we calculate the observed or sample value F_0 *

$$F_0 = 4S_z^2/3S_u^2 \tag{2-102}$$

and refer it to the table of percentage points of F, concluding that $\sigma_{e_3} < \sigma_e$, $\sigma_{e_3} = \sigma_e$, or $\sigma_{e_3} > \sigma_e$, depending on whether F_0 fell below the lower percentage point of F, or F_0 fell between the lower and upper percentage points of F, or F_0 fell above the upper percentage point of F, respectively.

For the case where $\sigma_{e_1} \neq \sigma_{e_2}$, but they are known accurately, see Ref. 12, p. 65, for significance tests and confidence bounds.

2-8.5.4 Shukla's Three-Instrument Bounds for $\sigma_{e_2}/\sigma_{e_1}$

Shukla (Ref. 15), apparently motivated by the paper of Hahn and Nelson (Ref. 18), who used one instrument twice, generalized their theory and extended the work of Grubbs in Ref. 12. Thus Shukla (Ref. 15) regarded the Hahn and Nelson (Ref. 18) approach as a special case of three independent instrument measurements (as does Grubbs in Refs. 2 and 12) and proceeds as follows. In fact, Shukla (Ref. 15) defines and uses

$$u_i = r_i - s_i \tag{2-103}$$

$$v_i = s_i - t_i \tag{2-104}$$

$$\delta = \sigma_{e_1}^2 / \sigma_{e_2}^2 \ (= \text{our } 1/k^2) \tag{2-105}$$

$$P = t_{1-\alpha}^2 / (t_{1-\alpha}^2 + n - 2) \tag{2-106}$$

*If $\sigma_{e_3}^2 = \sigma_e^2$, the quantity v of Eq. 2-92 equals 3/4.

$$A = r_{uv}^2 - P \tag{2-107}$$

$$B = 2[(r_{uv}^2 - P) + (1 - P)S_{uv}/S_v^2] \tag{2-108}$$

$$C = r_{uv}^2 - P + (1 - P)\,[(S_u^2/S_v^2) + (2S_{uv}/S_v^2)] \tag{2-109}$$

where

$t_{1-\alpha}$ = upper α probability level of Student's t of u and v
S_{uv} = sample covariance of u and v
S_v^2 = sample variance of $v = s - t$
S_u^2 = sample variance of $u = r - s$.

With these defined quantities, Shukla (Ref. 15) then points out that the $(1 - 2\alpha)$ confidence bounds on $\delta = \sigma_{e_1}^2/\sigma_{e_2}^2$ are determined from

$$Pr[\delta_L < \delta < \delta_U] = 1 - 2\alpha \tag{2-110}$$

where the lower δ_L and upper δ_U confidence bounds are found from

$$[\delta_L, \delta_U] = [-B \pm (B^2 - 4AC)^{1/2}]/(2A). \tag{2-111}$$

Apparently, Shukla's confidence bounds given by Eq. 2-111 are much narrower than those of Hahn and Nelson (Ref. 18) as demonstrated by Shukla with the Hahn and Nelson sample data.

Of course, an obvious rotation of the subscripts will give confidence bounds on $\sigma_{e_2}^2/\sigma_{e_3}^2$ and $\sigma_{e_3}^2/\sigma_{e_1}^2$.

Actually, the basic models described herein are of much more general use than might appear at first. Readers will, in general, have much familiarity with least squares and regression (Chapter 6) and thus will perhaps have experienced the analysis of residuals about a fitted curve. There may be some relation between standard error of residuals and our imprecision of measurement sigma. Moreover, if several instruments are used to take the same basic physical data and their residuals properly "paired", the techniques of this chapter may still apply. Thus once a satisfactory model or curve has been fitted, an analysis of the imprecision and inaccuracy of measurement can be made on the "residuals" or "errors of measurement".

We will illustrate Shukla's three-instrument method (Ref. 15) for I_2 and I_3 of Table 2-4. We "advance the subscripts" and calculate

$$S_v^2 = 0.2711, \qquad S_w^2 = 0.2252, \qquad r_{vw} = -0.8847, \qquad S_{vw} = -0.2186$$

$P = 0.3317$ from Eq. 2-106 $\alpha = 0.025$; and $A = 0.4510$, $B = -0.3954$,

$C = -0.0419$ from Eqs. 2-107, 2-108, and 2-109, respectively.

Finally, from Eq. 2-111

$$\delta_L = -0.0956, \quad \delta_U = 0.97.$$

Hence

$$Pr[0 \le \sigma_{e_2}^2/\sigma_{e_3}^2 < 0.97] = Pr[0 \le \sigma_{e_2}/\sigma_{e_3} \le 0.98] = 0.95.$$

(Had we calculated lower and upper 95% confidence bounds on $\sigma_{e_1}^2/\sigma_{e_2}^2$ using Shukla's method, both bounds would have been negative, due perhaps to $\sigma_{e_3}^2$!)

2-9 MEASUREMENTS WITH A GENERAL NUMBER $N \geq 3$ OF INSTRUMENTS

The separation of product variability and instrumental imprecision for any general number of measuring instruments was investigated in 1948 by Grubbs (Ref. 2) and later in 1964 by Jaech (Ref. 19). We will

define e_{ij} as before to be the random error of measurement for the ith reading by the jth instrument ($j = 3, 4, \ldots, N$), which measures the true unknown quantities x_i ($i = 1, 2, \ldots, n$), which may vary randomly or even be constant. If we use the notation of Ref. 2 where

$$S^2_{x+ej} = \text{sample variance in readings of the } j\text{th instrument } \mathrm{I}_j$$

$$S_{x+e_j,\ x+e_k} = \text{sample covariance of the sum of readings of instruments } \mathrm{I}_j \text{ and } \mathrm{I}_k$$

$$S^2_{e_j-e_k} = \text{sample variance of the difference in readings of instruments } \mathrm{I}_j \text{ and } \mathrm{I}_k,$$

the best estimate of the variance of errors of measurement of the first instrument I_1 for $N \geq 3$ is

$$\begin{aligned}\text{est}\sigma^2_{e_1} &= S^2_{x+e_1} - \left(\frac{2}{N-1}\right)\sum_{j=2}^{N} S_{x+e_1,\ x+e_j} + \left[\frac{2}{(N-1)(N-2)}\right]\sum_{2\leq j<k}^{k=N} S_{x+e_j,\ x+e_k} \\ &= \left(\frac{1}{N-1}\right)\left[\sum_{j=2}^{n} S^2_{e_1-e_j} - \left(\frac{1}{N-2}\right)\sum_{2\leq j<k}^{k=N} S^2_{e_j-e_k}\right]. \end{aligned} \tag{2-112}$$

The variance of the estimate given by Eq. 2-112 for normally distributed errors is

$$\text{Var}(\text{est}\sigma^2_{e_1}) = \left(\frac{2}{n-1}\right)\sigma^4_{e_1} + \left(\frac{1}{n-1}\right)\left\{\left[\frac{4}{(N-1)^2}\right]\sum_{j=1}^{N}\sigma^2_{e_1}\sigma^2_{e_j} + \left[\frac{4}{(N-1)^2(N-2)^2}\right]\sum_{2\leq j<k}^{k=N}\sigma^2_{e_j}\sigma^2_{e_k}\right\} \tag{2-113}$$

Formulas for estimates of $\sigma^2_{e_2}, \sigma^2_{e_3}, \ldots \sigma^2_{e_N}$ and the variances of these estimates may be found by rotation of the subscripts. In fact, one may merely designate the instrument he is interested in or working with as I_1 and use Eqs. 2-112 and 2-113 repeatedly until all instruments are covered.

The estimate of product variance of $N \geq 3$ instruments is the average of all of the sample covariances or

$$\text{est}\sigma^2_x = \left[\frac{2}{N(N-1)}\right]\sum_{1\leq j<k}^{k=N} S_{x+e_j,\ x+e_k} \tag{2-114}$$

$$= S^2_{[x+(e_1+\cdots+e_N)/N]} - \frac{1}{N^2(n-1)}\sum_{1\leq j<k}^{k=N} S^2_{e_j-e_k}$$

where the subscript $[x + (e_1 + \cdots + e_N)/N]$ means the average of the readings of all N instruments for the ith (and other) items(s).

The variance of the product variability estimate (Eq. 2-114) for normally distributed variables is

$$\text{Var}(\text{est}\sigma^2_x) = \left(\frac{2}{n-1}\right)\sigma^2_x + \left(\frac{1}{n-1}\right)\left\{\left(\frac{4}{N^2}\right)\sigma^2_x\sum_{j=1}^{n}\sigma^2_{e_j} + \left[\frac{4}{N^2(N-1)^2}\right]\sum_{1\leq j<k}^{k=N}\sigma^2_{e_j}\sigma^2_{e_k}\right\}. \tag{2-115}$$

In 1964 Jaech (Ref. 19) studied a measurement error model for the case where readings of N instruments are recorded on n items but where also r_j "runs" are made on instrument I_j ($j = 1, 2, \ldots, N$). Since the total number of data points is then nR, where

$$R = \sum_{j=1}^{N} r_j \tag{2-116}$$

some particular "unscrambling" of the measurement errors is clearly necessary.

The model considered by Jaech (Ref. 19) is linear, with constants α_k and β_k to be determined and quantities e_{ik} representing the random error of measurement on the ith item and kth run, and is given by

$$r_{ik} = \alpha_k + \beta_k x_i + e_{ik} \tag{2-117}$$

where

r_{ik} = observed value or reading on ith item for "run" k
i = 1, 2, . . ., n (i refers to ith item)
k = 1, 2, . . ., R (R = total number of readings)
x_i = true value of ith item measured.

In Jaech's model the parameters α_k and β_k are "joint" measures of instrument bias for "run" k. In fact, if $\alpha_k = 0$ and $\beta_k = 1$, no bias exists, but if $\beta_k = 1$, and $\alpha_k \neq 0$, there is a constant bias for the instrument on run, and the bias is independent of the magnitude of the measured item. Moreover, the possibility that $\beta \neq 1$ is not often considered in most applications. All unknown parameters in the model can be estimated by using sample covariances S_{jk} and variances S_k^2, as shown in Ref. 19, and are

$$\hat{\beta}_k = \left(\prod_{j \neq 1,k}^{R} S_{jk}/S_{1j} \right)^{1/(R-2)}, \quad k \neq 1 \tag{2-118}$$

$$\hat{\sigma}_x^2 = \left(\prod_{\substack{k=2 \\ k<q}}^{R} S_{1k}S_{1q}/S_{kq} \right)^{2/[(R-1)\,(R-2)]} \tag{2-119}$$

$$\hat{\sigma}_{e1}^2 = S_1^2 - \hat{\sigma}_x^2 \tag{2-120}$$

$$\hat{\sigma}_{ek}^2 = S_k^2 - \hat{\beta}_k \hat{\sigma}_x^2, \quad k \neq 1 \tag{2-121}$$

$$\hat{\alpha}_k = \bar{r}_k - \hat{\beta}_k \bar{r}_1, \quad k \neq 1 \tag{2-122}$$

$$\hat{\mu}_x = \bar{r}_1 \text{ (estimate of mean } x) \tag{2-123}$$

where

$\hat{}$ = estimate of quantity under the $\hat{}$
$\bar{r}_k$ = mean of readings on kth "run"
$\bar{r}_1$ = mean of readings on run 1.

As indicated in Jaech's paper (Ref. 19), the "run" designated as 1 is chosen as the base run, and therefore, for example,

$\hat{\beta}_k$ actually estimates β_k/β_1

and

$\hat{\alpha}_k$ actually estimates $\alpha_k - \beta_k\alpha_1/\beta_1$.

The relative biases between runs are independent of the base chosen although the estimate $\hat{\mu}_x$ of the mean product value and the estimate $\hat{\sigma}_x^2$ of product variance do depend on the base run, but normally they are only of interest in solving for estimates of the other parameters, i.e., the imprecisions.

Jaech (Ref. 19) also gives expressions for variances and covariances of the estimators and methods of comparison including an analysis of variance. In another paper Jaech (Ref. 20) extends this research

investigation to develop large sample tests of various hypotheses on instrumental precision for more than two instruments. It is evident that there should be many applications for the models studied by Jaech.

A computer program for estimating precision of measurement in accordance with the models of Ref. 2 or Eqs. 2-112 through 2-115 for any number of instruments has been written, thoroughly checked, and applied to various problems by O'Bryon (Ref. 21).

2-10 INTERLABORATORY TESTING FOR PRECISION AND ACCURACY STUDIES

One of the very important, practical, current, and ever-continuing problems in studies of precision and accuracy of measurement is that of interlaboratory testing. In this connection, it has become common practice to send "standard" or "reference" material to a number of laboratories for testing in order that analyses of the goodness of laboratory measurements can be established. Also it is desired to "bring the different laboratories into line" by providing calibrations. The standard or reference material tested at a number of laboratories is selected to be of consistent quality, very small variation if possible, or otherwise "homogeneous". In this way, the differences arising during the "round-robin" tests of the material at the different laboratories will reflect primarily, or hopefully, the differences in errors of measurement among the testers. However, there is bound to be some variation in the material tested that is not ordinarily stripped out of the laboratory instrument readings, as we have done previously in the chapter, to get at an analysis of only the errors of measurement. In addition, one has to be on guard in interlaboratory testing for "outliers", which nearly always arise because there may have to be some treatment or elimination of spurious readings or observations.

The precision of measurement at one (a single) laboratory will ordinarily be measured in terms of the standard deviation or variance in errors of measurement and is widely referred to as the "repeatability" sigma or value. Some will contend that repeatability should be measured in terms of a single operator on a single piece of measuring equipment at a single laboratory. We will avoid such arguments because it becomes most natural to identify, take into account, and estimate *all* of the components of variation that might arise in any particular problem facing the analyst or statistician.

The variation among the true levels or large sample average readings of the laboratories at which the round-robin procedure is conducted, when compared with the repeatability, is rather widely referred to as the "reproducibility" sigma or value. The reproducibility sigma involves not only the variation among true (or large sample) averages of the readings at each laboratory but also depends on the repeatability sigma of a laboratory — and, indeed, the number of measurements taken at a laboratory! In our example that follows we will make specific calculations and precise estimates of the components of variance involved and will illustrate the procedure in all necessary detail.

Although it is now often customary to include a fairly large number of laboratories (even 30 or 40) in a round-robin test, we will illustrate the problem for only seven laboratories since this will suffice for making our primary points.

Our illustration of the problem of interlaboratory testing consists of the determination by each of seven laboratories of the amount of lead in standard samples of gasoline. The particular samples of gasoline made up for the purpose of interlaboratory testing contained precisely 0.029 g/gal., and either two or three measurements or determinations (duplicate or triplicate) were recorded at each of the seven laboratories in the round-robin procedure. The data, taken from Ref. 22, on the measurements of the amount of lead in standard gasoline samples are given on Table 2-7, where the determined amounts of lead have been multiplied by 1000 for convenience of analysis.

There are a total of $N = 17$ measurements for all seven laboratories, and we define the following symbols for our use here:

x_{ij} = element (determined amount of lead in gasoline $\times$ 1000) or observation in the ith row and jth column of Table 2-7

$\Sigma x = \Sigma\Sigma x_{ij}$ = sum of all the observations in Table 2-7

$\Sigma x^2 = \Sigma\Sigma x_{ij}^2$ = sum of squares (SS) of all the observations in Table 2-7

$(\Sigma x)^2/N$ = Table 2-7 total squared divided by N = the "correction term"

n_j = number of observations in the jth column = 2 or 3 for Table 2-7

k = number of laboratories participating = 7

σ_r = repeatability sigma, or standard deviation, within laboratories

σ_L = standard deviation among true laboratory means or levels, or "external" sigma

$\sigma_R = \sqrt{\sigma_L^2 + \sigma_r^2}$ = reproducibility sigma for a single observation at a laboratory. (2-124)

The reader with some statistical background will recognize the data of Table 2-7 as a standard one-way classification in the analysis of variance (ANOVA) with an unequal number of observations per cell. The method of statistical analysis is given directly in Tables 2-7 and 2-8 and may be found in many standard textbooks on statistics.

Since there are unequal numbers of observations per cell in Table 2-7, some care must be exercised in estimating the components of variance, as we will see.

The numerical ANOVA is summarized in Table 2-8. There are a total of 16 df, with 10 for the residual or repeatability variance σ_r^2, and the remaining 6 df are equal to one less than the number 7 of laboratories.

TABLE 2-7

ONE-WAY ANOVA CLASSIFICATION FOR LEAD IN GASOLINE
(0.029 LEVEL; VALUES MULTIPLIED BY 1000)

DuPont	Mobil	EPA	Ethyl	Amoco	Ford	Octel
23	24	25	26	28	27	28
24	24	26	26	27	27	28
23	...	26	...	...	26	...
70	48	77	52	55	80	56

$N = 17$, x_{ij} = element in ith row and jth column

$$\Sigma x = \Sigma\Sigma x_{ij}^2 = 23 + 24 + 23 + 24 + 24 + \cdots + 28 + 28 = 438$$

$$\Sigma x^2 = \Sigma\Sigma x_{ij}^2 = (23)^2 + (24)^2 + (23)^2 + (24)^2 + (24)^2 + \cdots + (28)^2 + (28)^2 = 11{,}330$$

$$(\Sigma x)^2/N = (438)^2/17 = 11{,}284.94$$

$$\text{Total SS (about grand mean)} = \Sigma x^2 - (\Sigma x)^2/N = 11{,}300 - 11{,}284.94 = 45.06$$

$$\text{SS among column (Lab) means} = (70)^2/3 + (48)^2/2 + (77)^2/3 + \cdots + (56)^2/2 - 11{,}284.94$$
$$= 11{,}327.50 - 11{,}284.94 = 42.56$$

$$\text{SS for repeatability within Labs} = 45.06 - 42.56 = 2.50.$$

Copyright, American Society for Testing and Materials, 1916 Race Street, Philadelphia, PA 19103. Reprinted with permission.

TABLE 2-8

ANOVA TABLE

Source of Variation	Sum of Squares	df	Variance
Total	45.06	16	
Among Labs	42.50	6	$7.093 = \sigma_r^2 + 2.41\,\sigma_L^2$
Within Labs	2.50	10	$0.25 = \sigma_r^2$

$2.41 = (N^2 - \Sigma n_j^2)/[N(k-1)]$, $\sigma_r = 0.50$, $\sigma_L = 1.69$, and $\sigma_R = \sqrt{\sigma_L^2 + \sigma_r^2} = 1.76$

Copyright, American Society for Testing and Materials, 1916 Race Street, Philadelphia, PA 19103. Reprinted with permission.

The residual or repeatability variance σ_r^2 is rather small; the estimate of it is

$$\hat{\sigma}_r^2 = 2.50/10 = 0.25 \quad \text{or } \hat{\sigma}_r = 0.5$$

which converts to only $0.5/1000 = 0.0005$ in g/gal. of lead.

Note that the variation among laboratory true levels of measurement is quite large and highly significant with

$$F = 7.093/0.25 = 28.4$$

whereas $F_{0.99}$ (6,10) is only 5.36. We must conclude, therefore, that the component of variance among laboratory measurement levels is rather large and deserves investigation to "bring the laboratories into line" by providing calibration corrections.

To estimate the component of variance among laboratory true levels of measurement, we must equate

$$\hat{\sigma}_r^2 + \left[\frac{N^2 - \Sigma n_j^2}{N(k-1)}\right]\hat{\sigma}_L^2 = \hat{\sigma}_r^2 + 2.41\hat{\sigma}_L^2 = 7.093 \tag{2-125}$$

from which we obtain

$$\hat{\sigma}_L^2 = 2.84 \quad \text{or} \quad \hat{\sigma}_L = 1.69 \ \ (0.00169 \text{ g/gal.}).$$

Finally, the reproducibility variance σ_R^2 for n measurements at a laboratory taken at random would be

$$\sigma_R^2 = \sigma_L^2 + \sigma_r^2/n = 2.84 + 0.25/n. \tag{2-126}$$

For the average result of k laboratories, Eq. 2-126 would be divided by k, the number of laboratories.

We will not discuss "outlying" laboratories in this chapter since "outliers" are covered in Chapter 3. Our prime interest is to show how the analysis should be conducted without rejecting any laboratory results at this stage.

With reference to the interlaboratory test one notes that each and every measurement of the amount of lead in gasoline is consistently lower than the actual amount, i.e., 0.029 g/gal.; thus all laboratories show low readings. Some calibration is necessary, especially after some investigation to determine the possible cause of the consistently low measurements. In fact, by examining Table 2-9, we see that DuPont and Octel differ by $28.0 - 23.3 = 4.7$, which is $4.7/1.76 = 2.7$ times the reproducibility sigma of a single

TABLE 2-9

AVERAGE LEVELS OF THE DIFFERENT LABORATORIES

DuPont	Mobil	EPA	Ethyl	Amoco	Ford	Octel
23.3*	24.0*	25.7	26.0	27.5	26.7	28.0

*The levels of measurement at DuPont and Mobil appear significantly lower than the other laboratories.

measurement! Apparently, there is no problem concerning the within-laboratory or repeatability sigma of 0.50, but the laboratories urgently need bringing into line by calibration for average readings.

Finally, we caution again that in this type of interlaboratory analysis of a test program, we are not necessarily dealing strictly with the errors of measurement to determine precision and accuracy as previously stripped out as components in this chapter. We say this, even though in this particular round-robin test there may be little, if any, variation due to the product, i.e., amount of lead. It can often be expected, nevertheless, that some product variation may still be present in ordinary interlaboratory testing even though it would be highly desirable to deal only with errors of measurement for precision and accuracy studies of a test method as we have presented and recommended predominantly.

The reader should note in particular that the interlaboratory test and the multi-instrument cases discussed heretofore can sometimes, and often should, be treated as the same analytical procedure. In fact, the multi-instrument analysis seems more general.

2-11 THE HIERARCHY OF CALIBRATIONS AND THE ACCUMULATION OF ERROR

As the final major topic to be highlighted in this chapter, we believe it pertinent to discuss the problem of calibration of instruments up through the various calibration echelons to the prime reference standards at the National Bureau of Standards and also to discuss the accumulation of error throughout the chain. We have seen that both precision and accuracy are very important or mandatory, that instrumental precision is required to detect bias or systematic error, and that bias or improper levels of measurement may be corrected by good calibration or bias correction procedures.

Crow (Ref. 23) gives a brief account of the background of the calibration process, which will suffice for our needs in this chapter. We quote Crow (Ref. 23).

"Since the art of measurement began there have been standards, more or less informal, by means of which further measuring sticks, weights, and capacity measures have been produced for use in construction and commerce. With each reproduction of the measures variations were inevitably introduced, and these often consisted of intentional as well as accidental errors. The ancient Egyptians, Greeks, and Romans had respected standards of measure, but these fell out of use during the Dark Ages, and the later attempts to establish widely used standards were long doomed to failure.

"In 1830 the United States Senate noted that variations in the standards in use at various customhouses were causing loss of revenue and directed the Secretary of the Treasury to make comparisons of these standards. The Treasury in fact took steps to supply uniform weights and measures to all customhouses, and the Secretary reported in 1832 that standards were being fabricated at the United States Arsenal in Washington 'with all the exactness that the present advanced state of science and the arts will afford'. Thus the Office of Weights and Measures came to be established in the late 1830's within the Treasury Department. In 1901, when its budget was still less than $10,000, the Office became a part of the new National Bureau of Standards. In 1903 the Bureau was transferred to its present position in the Department of Commerce.

"Now the Bureau maintains hundreds of national standards and calibrates the standards of the states, military departments, manufacturers, utilities, universities, private testing companies, and others. The Bureau is unable to calibrate all secondary standards and instruments, and the above types of organizations in turn calibrate further standards. For example, counties and cities may have their balances,

weights, and other measures certified by their state offices, and they in turn certify the balances within their jurisdictions.

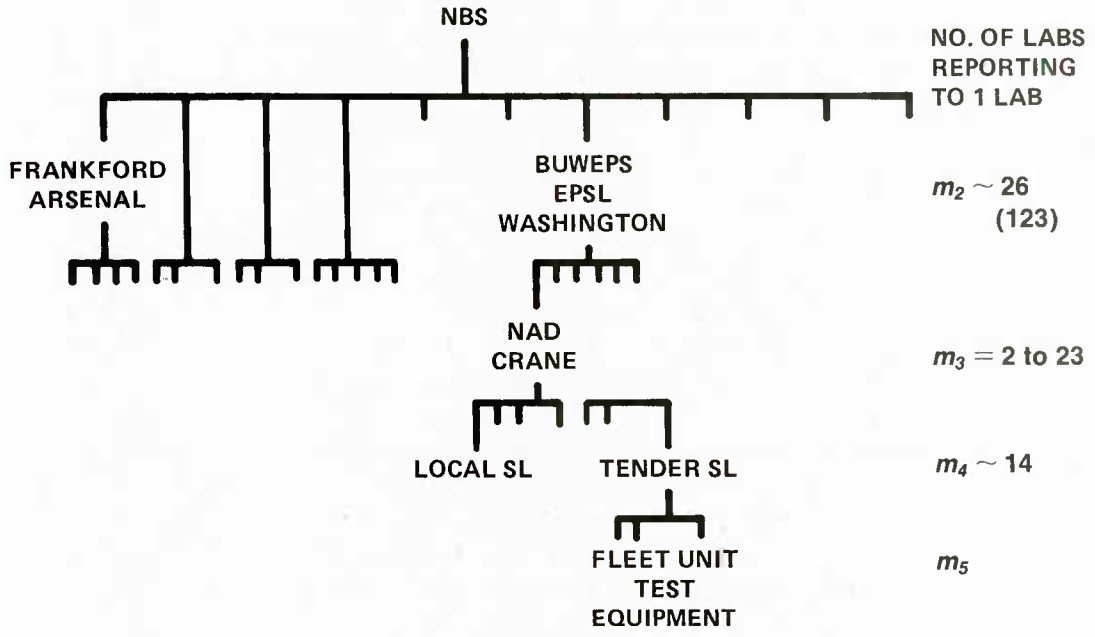
"In electrical energy the Bureau uses a standard watthour meter accurate to about 0.03 percent to calibrate the master standards of public utility commissions and power companies. The latter in turn make measurements to about 0.1 percent of customers' meters. As a result in part of variability in time, customers' meters operate within about one percent accuracy.

"In recent years the demanding requirements of missiles, spacecraft, and other vehicles have led to the establishment of extensive hierarchies of standards laboratories by the military departments. As indicated in Fig. 1 [our Fig. 2-1], the National Bureau of Standards is at the apex of these hierarchies. The figure indicates just a few examples of the standards laboratories that enter in various levels, or echelons, of the hierarchy. For most basic standards the Bureau is itself just one of the many national laboratories deriving their units from the International Bureau of Weights and Measures. In each echelon of the hierarchy and with each transfer of information, some error is unavoidably introduced."

With this coverage of the calibration process, let us take a brief look at the need for precision of measurement for each level at which the instrument may be calibrated and used for measurement purposes along with the accumulation of error in the instrument comparison process. We will number the echelons at which calibrations may occur with the numbers 1, 2, . . . , m, where the first level or 1 refers to the National Bureau of Standards, 2 the second level, and so on down to the final laboratory or "bench" level m where measurements are taken on some item. Then at each and every level or echelon an error in calibration may be committed, or that is, we may say that the error committed at level i is e_i. Hence if the calibrations at the different echelons are statistically independent, as we would expect, the total variance σ_T^2 of the errors down to the mth level is

$$\sigma_T^2 = \sum_{i=1}^{m} \sigma_{e_i}^2 = m\sigma_e^2 \tag{2-127}$$

if the same standard error σ_e of measurement is made at each level. It might be expected, however, that precision of measurement should improve as the numbered level decreases, i.e., 5, 4, 3, 2, and 1. Thus the number m of levels may be of some importance although the relative precision in measuring product



Reprinted with permission. Copyright © by American Society for Quality Control.

Figure 2-1. Schematic Representation of Hierarchies of Military Standards Laboratories Using National Bureau of Standards Calibration Services (Ref. 23)

variability is of considerably more importance. In fact, to demonstrate this, recall that actual measurements will be made at the mth or last level, so that with the hope that no confusion will arise, one may take

$$\sigma_{m+1} = \sigma_x, \text{ when } i \text{ becomes } m, \tag{2-128}$$

i.e., the $(m + 1)$ st sigma is the actual product standard deviation measured. What is important then is really the precision ratio r_i (often misnamed the accuracy ratio)

$$r_i = \sigma_x / \sigma_{e_i} \tag{2-129}$$

at each level, and the accumulated variance (Eq. 2-127) at level m.

The accumulation of calibration error or variance of the errors throughout the hierarchy of calibration echelons has been studied very thoroughly by Woods and Zehna (Ref. 24) and particularly also in cost or economic detail by E. Crow (Ref. 23).

It seems reasonable to define the resultant or total precision or accuracy ratio r_T, say, as

$$r_T^2 = \sigma_x^2 / \sum_{i=1}^{m} \sigma_{e_i}^2 \tag{2-130}$$

where total accumulation of variances in errors of measurement are accounted for and included. If at each stage i the relative precision ratio is constant, i.e.,

$$r_i = \sigma_x / \sigma_{e_i} = \text{c} \tag{2-131}$$

Woods and Zehna (Ref. 24) have shown that the final or total precision ratio (Eq. 2-130) is simply

$$r_T^2 = \frac{(c^2 - 1)c^{2m}}{c^{2m} - 1} . \tag{2-132}$$

As the number m of echelons of calibration increases without limit, r_T^2 approaches

$$\lim_{m \to \infty} r_T^2 = c^2 - 1. \tag{2-133}$$

Thus always

$$r_T^2 \geq c^2 - 1 \tag{2-134}$$

and r_T never falls below

$$r_T = \sqrt{c^2 - 1} \tag{2-135}$$

which is a very enlightening result indeed! Hence as a numerical example, if we require

$$r_i = \sigma_x / \sigma_e = 10$$

the total precision or accuracy ratio does not fall below

$$r_T = \sqrt{(10)^2 - 1} = 9.95!$$

Crow (Ref. 23) shows that if at each calibration stage

$$\sigma_{i+1} / \sigma_i \geq 2$$

for a large number of calibration echelons, the relative total precision σ_T/σ_m will not increase by more than about 15%. Thus as $m \to \infty$, it is only when σ_{i+1}/σ_i becomes less than 2 that one should expect any very significant or intolerable accumulation of relative total calibration error precision σ_T/σ_m.*

Crow (Ref. 23) conducts a very fine study of the optimum allocation of calibration errors based on total system cost of achieving a given or desired accuracy. He considers costs to be of two types: (1) the cost of research and development (R&D) that needs to be done only once or not at all if the measurement system has already been developed, and (2) the costs of installation and operation for each laboratory. Crow (Ref. 23) then assumes that both types of costs decrease in a negative exponential manner with increasing size of the error E committed in a laboratory, i.e.,

$$\text{R\&D cost} \approx b_0 E^{-a_0} \tag{2-136}$$

and

$$\text{Installation and Operation Cost} \approx b_1 E^{-a_1} \tag{2-137}$$

where all constants are positive and

$$a_0,\ b_0,\ a_1,\ b_1 = \text{fitted constants with } a_0 \geq a_1.$$

By using the method of Lagrange Multipliers to minimize total costs, Crow (Ref. 23) finds that the optimum precision error ratio between the ith and $(i+1)$st stage of the calibration echelons is given by

$$\sigma_{i+1}/\sigma_i = (m_{i+1})^{1/(a+2)} \tag{2-138}$$

where

$$m_{i+1} = \text{number laboratories at stage } i+1$$

and $0 < a \leq a_1$, and $a = a_1$ if research and development is unnecessary. Hence the exponent value a becomes equal to a_1, or the exponent of the installation and operating cost curve, if no R&D is required for the instrumentation.

2-12 ADDITIONAL DISCUSSION OF FUNDAMENTALS OF MEASUREMENT

The American Society for Testing and Materials (ASTM) has published (1977) a compendium of standards on precision and accuracy (Ref. 26). It is referred to as their "Green book" and may be of some interest to readers especially concerning just how precision and accuracy problems are now handled in much industrial work or practice. ASTM also has a standard recommended practice, designated E 177-71, entitled *Use of the Terms Precision and Accuracy as Applied to Measurements of a Property of a Material*, which may be found in the "Green book" (Ref. 26), pp. 124-41.

As indicated earlier in the chapter, a rather informative and thorough discussion of the precision and accuracy problem in many areas of the physical sciences is covered in Ref. 1. Also concerning the precision and accuracy of the fundamental constants in physics and the needed adjustment of them, the reader is referred to Eisenhart (Ref. 27) in addition to the many papers in Ref. 1.

Pontius discusses the fundamentals of measurement and the consideration of measurement as a production process in Ref. 28.

Cameron (Ref. 29) discusses the general problem of measurement assurance, and DeVoe (Ref. 30) examines the area of validation of the measurement process.

Mandel (Ref. 31) discusses the measurement process, especially in terms of interlaboratory testing.

The Engineering Design Handbooks (Refs. 32, 33, 34, 35, and 36) on experimental statistics constitute a very useful background of statistical knowledge for the reader concerning this chapter and also the other chapters of this handbook.

Finally, we comment on some very recent accomplishments concerning the three-instrument case, which should have wide applications. As is evident from Eq. 4-2, the models represented by Eq. 2-15 and Eq.

*The effect of calibration on end-item performance in echelon systems is discussed and modeled in Hilliard and Miller (Ref. 25).

6A-1 of the Appendix 6A, the estimation techniques for the imprecision of measurement are very closely tied in with the two-way ANOVA concept. Indeed, in a private communication Professor Ralph Bradley and Dennis Brindley (1980) of Florida State University indicate some very striking results for the three-instrument ($j = 3$) case. They use r_{ij}, which we have designated in this chapter to be the ith reading of the jth instrument, to mean the element of a two-way classification of the ith row and jth column. Thus as in the analysis of variance modeling, the sum of the instrumental biases β_j can be taken to be zero (but are still representative) and the variance $\text{Var}(e_{ij}) = \sigma^2_{e_j}$. Then, upon taking

$$S_j = \sum_{i=1}^{n} (r_{ij} - \bar{r}_{i\cdot} - \bar{r}_{\cdot j} + \bar{r}_{\cdot\cdot})^2 \tag{2-139}$$

where the dots simply denote summing on that particular subscript and the bars average values, and using the quantities

$$Q_j = kS_j/[(n-1)(k-2)] - \sum_{j=1}^{k} S_j/[(n-1)(k-1)(k-2)] \tag{2-140}$$

one finds for $k =$ the number of columns (instruments in this chapter) that the expected value of Q_j is

$$E(Q_j) = \sigma^2_{e_j} \tag{2-141}$$

our imprecision variance of measurement for the jth instrument, or here the residual variance in the jth column when row and column level effects have been eliminated, leaving "measurement errors". For the case $k = 3$, Brindley and Bradley indicate they have found the joint probability density of the Q_1, Q_2, and Q_3 and have established the likelihood ratio test of the null hypothesis

$$H_0\colon \sigma^2_{e_1} = \sigma^2_{e_2} = \sigma^2_{e_3} = \sigma^2_e \tag{2-142}$$

versus the alternative hypothesis

$$H_a\colon \text{Some } \sigma^2_{e_j} \neq \sigma^2_{e_q},\ j \neq q. \tag{2-143}$$

The likelihood ratio statistic for testing H_0 is

$$\lambda = 3(Q_1Q_2 + Q_1Q_3 + Q_2Q_3)/(Q_1 + Q_2 + Q_3)^2 \tag{2-144}$$

and under H_0 the probability density of λ is simply

$$f(\lambda) = \left(\frac{n-2}{2}\right)\lambda^{(n-4)/2},\ 0 \leq \lambda \leq 1 \tag{2-145}$$

so that any α probability level of λ, or λ_α, will be given by

$$\lambda_\alpha = (\alpha)^{2/(n-2)}. \tag{2-146}$$

Brindley and Bradley have also established the power function of the test of H_0 for the case of $k = 3$ instruments.

2-13 SUMMARY

We have defined errors of measurement and the terms precision and accuracy of measurement in rather extensive and analytical detail, approaching the problem primarily from the practical point of view of requirements. Methods and techniques for estimating precision and accuracy of measurement for various

numbers of instruments used in the process are thoroughly covered along with statistical tests of significance on the parameters of imprecision and inaccuracy, and confidence bounds as well. Related work of many authors on the problem of precision and accuracy is discussed, and references to industrial practice are given. Finally, we present an account of the hierarchy of calibrations for instruments and indicate precision requirements for each echelon of laboratory calibrations.

Many examples concerning applications of the currently available theory of precision and accuracy are presented throughout to orient the reader as well as possible.

The methods of this chapter are especially recommended to accumulate data on precision, accuracy or bias, and calibration corrections for all instruments in order that instrumental capabilities will be documented and appropriate selections of the best or standard reference instruments can be made as needed in the overall measurement process.

REFERENCES

1. National Bureau of Standards, *Precision Measurement and Calibration (Statistical Concepts and Procedures)*, Special Publication 300, **1**, US Department of Commerce, February 1969.
2. Frank E. Grubbs, "On Estimating Precision of Measuring Instruments and Product Variability", Journal of the American Statistical Association **43**, 243-64 (June 1948).
3. W. A. Thompson, Jr., "Precision of Simultaneous Measurement Procedures", Journal of the American Statistical Association **58**, 474-9 (June 1963).
4. R. Choudary Hanumara, "Estimating Imprecisions of Measuring Instruments", Technometrics **17**, 299-302 (August 1975).
5. You-Yen Yang, "Modification of Grubbs' Estimators When Errors of Measurement are Correlated", manuscript 1982 available from Mr. Yang at the US Environmental Protection Agency, Washington, DC 20460.
6. S. N. Roy and R. C. Bose, "Simultaneous Confidence Interval Estimation", Annals of Mathematical Statistics **24**, 513-36 (1953).
7. Norman Draper and Irwin Guttman, "*k* Simultaneous Measurement Procedures: A Bayesian Approach", Journal of the American Statistical Association **71**, 605-7 (September 1976).
8. E. J. G. Pitman, "A Test on Normal Correlation", Biometrika **31**, 9-12 (1939).
9. W. A. Morgan, "A Test for the Significance of the Difference Between Two Variances in a Sample from a Normal Bivariate Population", Biometrika **31**, 13-9 (1939).
10. C. J. Maloney and S. C. Rastogi, "Significance Tests for Grubbs' Estimators", Biometrika **26**, 671-6 (1970).
11. Frank E. Grubbs and James F. O'Bryon, *The Statistical Comparison of Measuring Instruments (A Study of Precision and Accuracy Problems for NATO Velocity Measuring Chronographs)*, US Army Aberdeen Research and Development Center Technical Report No. 11, Aberdeen Proving Ground, MD, July 1971.
12. Frank E. Grubbs, "Errors of Measurement, Precision, Accuracy and the Statistical Comparison of Measuring Instruments", Technometrics **15**, 53-66 (February 1973).
13. S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses", The Annals of Mathematical Statistics **9**, 60-2 (1938).
14. G. K. Shukla, "Some Exact Tests of Hypothesis About Grubbs' Estimators", Biometrics **29**, 373-7 (1973).
15. G. K. Shukla, "Confidence Intervals on Variance Ratios", (Letter to the Editor) The American Statistician **30**, 151-2 (August 1976).
16. W. A. Thompson, Jr., "Estimation of Dispersion Parameters", Journal of Research of the National Bureau of Standards—B. Mathematics and Mathematical Statistics **66B**, 161-4 (October-December 1962).
17. R. C. Hanumara and W. A. Thompson, Jr., "Percentage Points of the Extreme Roots of a Wishart Matrix", Biometrika **55**, 505-12 (November 1968).

18. J. H. Hahn and W. Nelson, "A Problem in the Statistical Comparison of Measuring Devices", Technometrics **12**, 95-102 (February 1970).
19. J. L. Jaech, "A Program to Estimate Measurement Error in Non Destructive Evaluation of Reactor Fuel Element Quality", Technometrics **6**, 293-300 (August 1964).
20. John L. Jaech, "Large Sample Tests of Grubbs' Estimators of Instrument Precision with More Than Two Tests", Technometrics **18**, 127-33 (May 1976).
21. James F. O'Bryon, "Estimation of Precision of Measurement and Product Variability: A Computer Model", BRL Report No. 2778, August 1977.
22. Frank E. Grubbs, "Application for Precision Data to Specifications for Petroleum Products", American Society for Testing and Materials Standardization News **2**, 18-21 and 49 (July 1974).
23. Edwin L. Crow, "Optimum Allocation of Calibration Errors", Industrial Quality Control **23**, 215-9 (November 1966).
24. W. Max Woods and Peter W. Zehna, "Cumulative Effect of Calibration Errors", Industrial Quality Control **22**, 411-2 (February 1966).
25. J. E. Hilliard and J. R. Miller III, "The Effect of Calibration on End-Item Performance in Echelon Systems", Journal of Quality Technology **12**, 61-70 (April 1980).
26. *Standards on Precision and Accuracy for Various Applications*, American Society for Testing Materials, First Edition, Philadelphia, PA, 1977.
27. Churchill Eisenhart, "Contributions to Panel Discussion on Adjustments of the Fundamental Constants, Precision Measurement and Fundamental Constants", *Proceedings of the International Conference held at the National Bureau of Standards*, Gaithersburg, MD, August 3-7, 1970, edited by D. N. Langanberg and B. N. Taylor. NBS Special Publication 343. Government Printing Office, Washington, DC 20402. Issued August 1971.
28. Paul E. Pontius, *Notes on the Fundamentals of Measurement and Measurement as a Production Process*, NBSIR 74-545, Optical Physics Division, Institute for Basic Standards, National Bureau of Standards, Washington, DC, September 1974.
29. J. M. Cameron, *Measurement Assurance*, NBSIR 77-1240, Institute for Basic Standards, National Bureau of Standards, Washington, DC, April 1977.
30. James R. Devoe, "Validation of the Measurement Process", *American Chemical Society Symposium Series 63*, American Chemical Society, Washington, DC, 1977.
31. John Mandel, "The Measuring Process", Technometrics **1**, 251-67 (August 1959).
32. AMCP 706-110, Engineering Design Handbook, *Experimental Statistics, Section 1, Basic Concepts and Analysis of Measurement Data,* December 1969.
33. AMCP 706-111, Engineering Design Handbook, *Experimental Statistics, Section 2, Analysis of Enumerative and Classificatory Data,* December 1969.
34. AMCP 706-112, Engineering Design Handbook, *Experimental Statistics, Section 3, Planning and Analysis of Comparative Experiments,* December 1969.
35. AMCP 706-113, Engineering Design Handbook, *Experimental Statistics, Section 4, Special Topics,* December 1969.
36. AMCP 706-114, Engineering Design Handbook, *Experimental Statistics, Section 5, Tables,* December 1969.

BIBLIOGRAPHY

F. N. David, *Tables of the Correlation Coefficient*, published by the Biometrika Office, University College, London, 1938.

R. A. Fisher, "Mathematical Probability in the Natural Sciences", Technometrics **1**, 21-9 (February 1959).

R. Choudary Hanumara, "Estimating Variances in Simultaneous Measurement Procedures" (Letter to the Editor) The American Statistician **29**, 108-9 (May 1975).

E. A. Maxwell, "Estimating Variances from One or Two Measurements on Each Sample", The American Statistician **28**, 96-7 (August 1974).

E. S. Pearson and H. O. Hartley, Editors, *Biometrika Tables for Statisticians*, Vol. I, Cambridge University Press, Cambridge, 1954.

W. A. Thompson, Jr., "The Problem of Negative Estimates of Variance Components", The Annals of Mathematical Statistics **33**, 273-89 (March 1962).

W. J. Youden, W. S. Conner, and N. C. Severa, "Measurements Made by Matching with Known Standards", Technometrics **1**, 101-9 (May 1959).

CHAPTER 3
PROCEDURES FOR DETECTING OUTLYING OBSERVATIONS

Statistical principles for screening observational data to detect irregular or outlying observations are discussed in appropriate detail and illustrated by examples. The best tests that have been found to be extensively used in practice are covered for the problem of detecting single or multiple outliers in samples. The principal tests include those for detecting whether the highest or the lowest observations are outliers, or the two highest or the two lowest, or the highest and lowest observations jointly come from different populations with shifted means or a change in the dispersion parameter. Moreover, the principles are extended to the problem of detecting more than two or many outliers in data, and the relation of outlier detection to tests of normality is presented.

3-0 LIST OF SYMBOLS

$a = \sum_{i=k+1}^{n-k} x_i/(n-2k) =$ trimmed mean of Rosner

$a_{n-i+1} =$ coefficient of the Wilk-Shapiro statistic

$B^* =$ Hawkins and Perold's studentized maximum statistic of Eqs. 3-62 and 3-63

$b^2 =$ Rosner's trimmed variance in Eq. 3-49

$\sqrt{b_1} =$ sample skewness coefficient of Eq. 3-56

$b_2 =$ sample kurtosis coefficient of Eq. 3-57

$d =$ maximum studentized statistic of Halperin, Greenhouse, Cornfield, and Zalokar in Eq. 3-61

$E( ) =$ expected or mean value of quantity in parentheses

$E_k =$ Tietjen-Moore ratio statistic given by Eq. 3-44

$E_k^* = (S_k^2 + U)/(S^2 + U) =$ Hawkins' outlier test statistic of Eq. 3-55

$F = F( ) =$ cumulative distribution function

$f( ) =$ probability density function of quantity in parentheses

$H_0 =$ null hypothesis

$k =$ number of "outliers" in Tietjen-Moore tests

$L =$ bound or limit

$L_k =$ Tietjen-Moore ratio statistic given by Eq. 3-46

$\max|\ \ | =$ maximum value of quantity inside | |

$N =$ total number of items in a finite population

$n =$ number of observations in the sample

$P =$ level of probability

$Pr[y<y_0] = F(y_0) =$ chance y is less than y_0

$p =$ fraction of the total sample size

$R_1 =$ Rosner's maximum ratio in Eq. 3-50

$R_2 =$ Rosner's second largest ratio in Eq. 3-51

$r =$ number less than N

$r_i = |x_i - \bar{x}| =$ absolute residuals used by Tietjen and Moore to determine their z_i's (par. 3-5.5.2)

$r_{ij} =$ Dixon's statistics for testing outliers (See Table 3-2 for all of Dixon's definitions used in this chapter.) For example, $r_{11} = (x_n - x_{n-1})/(x_n - x_1.)$

$S^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 =$ total sum of squares about sample mean for the entire sample

S_k^{2*} = Hawkins' inlier sum of squares based on unsuspected sample values

$S_{n,n-1}^2$ = sum of squares, omitting the two highest sample values x_{n-1} and x_n

$S_{1,2}^2 = \sum_{i=3}^{n}(x_i - \bar{x}_{1,2})^2$ = sum of squares, omitting the two lowest sample values

s = sample standard deviation based on $(n-1)$ degrees of freedom

$s' = \sqrt{\Sigma(x_i - \bar{x})^2/n} = \sqrt{(n-1)/n}\, s$ = sample standard deviation based on total sample size n

s^2 = sample variance based on $(n-1)$ degrees of freedom (See Eq. 3-2.)

s_ν = independent estimate of the standard deviation based on ν degrees of freedom

$T_n = (x_n - \bar{x})/s$ = statistic for testing whether the largest sample value x_n is too large

T_x', T_y' = values based on coordinates x,y

$T_1 = (\bar{x} - x_1)/s$ = statistic for testing x_1

T_1', T_n' = values of T_1 and T_n based on an independent s_ν with ν degrees of freedom in Eqs. 3-59 and 3-60

$T_{1\infty}', T_{n\infty}'$ = critical T-values in Eqs. 3-65 and 3-66 based on known population standard deviation σ

$t^{(1)}$ = largest signed value of t_i (See Eq. 3-13.)

$t_i = (x_i - \bar{x})/s'$ (See Eq. 3-11.)

U^2 = independent sum of squares used by Hawkins in Eq. 3-54

W = Wilk-Shapiro statistic of Eq. 3-65

$w = x_n - x_1$ = sample range

w/s = ratio of sample range to sample standard deviation. Sometimes called the "studentized" range, although studentization usually calls for an independent s in the denominator.

w_0 = limit (of integration) for the range w

w_3 = range or maximum dispersion of a sample of three observations, i.e., largest minus smallest values

x_i = ith ordered sample value in order of magnitude $x_1 \leq x_2 \leq x_i \leq \cdots \leq x_n$

x_n = largest sample value

x_1 = smallest sample value

$x^{(1)}$ = sample value making R_1 a maximum

$\bar{x} = \sum_{i=1}^{n} x_i/n$ = sample mean

$\bar{\bar{x}}$ = grand mean

$\bar{x}_k = \sum_{i=1}^{n-k} x_i/(n-k)$, Tietjen-Moore mean

$\bar{x}_{n,n-1} = \sum_{i=1}^{n-2} x_i/(n-2)$ = mean, omitting x_{n-1} and x_n

$\bar{x}_{1,2} = \sum_{i=3}^{n} x_i/(n-2)$ = mean, omitting x_1 and x_2

x_i' = ith observation or sample value in the order taken, the original sample being $x_1', x_2', \cdots, x_i', x_n'$

x', x'', x''' = Lieblein's sample of three observations, where x' and x'' are the two closest values

$|x_i - x_k|$ = absolute difference or positive value of the difference between any two sample values x_i and x_k

x,y = variables of integration, or variables, also coordinates

$y = (x' - x'')/(x_3 - x_1)$ = Lieblein's ratio in Eq. 3-26

y_0 = a limit

z_i = original observed x that is the ith closest to the sample mean $\bar{x}$

z_i = Tietjen-Moore designation for the original observations x_i, such that z_i is that particular x for which the r_i is the ith ordered (increasing) absolute residual

$\bar{z}$ = mean of the full sample = $\bar{x}$ also

$\bar{z}_k$ = Tietjen-Moore mean of the $(n-k)$ least extreme observations given by Eq. 3-45

α = probability level = 0.05, 0.01, etc.

α_{1-F} = percentage point as in Eq. 3-31

β = probability level

$\lambda_i(\beta)$ = level for Rosner's R_i

μ = population or universe mean

σ = population standard deviation

$\sigma(\ )$ = standard deviation of quantity in parentheses

$\hat{\sigma}_r^2$ = estimate of the within variance σ_r^2

χ_0 = limit (of integration) for chi

χ^2 = $\chi^2(\nu)$ = chi-square with ν degrees of freedom

3-1 INTRODUCTION

In Chapter 2 we covered the problem of taking measurements and trying to control or assure the quality of them by knowing the precision and accuracy of our measuring instruments. In fact, it becomes of utmost importance to have at hand the capability of any measuring instrument we use in applications because taking action in the presence of errors of measurement would lead to unwarranted results or even to a costly state of affairs. Hence the need exists to control errors of measurement in all experiments by continuing to collect information on the precision and accuracy of our measuring instruments. Indeed, this should be a daily activity because measurements are expensive and should be taken with care.

Once we can insure that our measurements are of high quality, we may proceed with confidence that our analyses of the data are correct, and we can depend on any action taken as a result thereof. Perhaps one of the most appropriate next steps is to examine the data we take or acquire for the presence of "outliers". In fact, one or more of the errors of measurement could be due to the existence of outlying observations (unusually large errors of measurement), and it is important to examine the data for such measurements. For example, suppose we take the same measurements with two different measuring instruments as indicated in Chapter 2. We might list the differences in readings of the two instruments for each item or characteristic measured, and if one or more of the differences are large, we would certainly like to investigate the cause and possibly determine which instrument was at fault. Moreover, even if we made no errors of measurement or screened them out, our observations may still contain some deviant values. Also we would like to be able to judge whether there could have been a shift in level, or perhaps increased dispersion, other causes worth looking for, or whether the deviant values are truly characteristic of the items under study. Hence we must be aware that our data will often have to be screened not only for errors of measurement, but for "outliers", or outlying observations, as well. The purpose of this chapter is to present methods for detecting outlying observations and for treating them in further analyses.

An outlying observation, or an "outlier", is one of the sample values that appears to deviate markedly from the other members of the sample in which it occurs. In this connection, the two possible alternatives that follow are of some primary interest to us:

1. An outlying observation may be merely an extreme manifestation of the random variability inherent in the data. If this is true, the values should be retained and processed in the same manner as the other observations in the sample.

2. On the other hand, an outlying observation may be the result of gross deviation from the prescribed experimental procedure or an error in calculating or recording the numerical value. In such cases, it may be desirable to undertake an investigation to determine the reason for the aberrant value. The observation may even eventually be rejected as a result of the investigation, though not necessarily so. At any rate, in subsequent data analysis the outlier or outliers will be recognized as probably being from a different population than that of the other sample values.

It is our purpose to provide statistical rules that will lead the experimenter almost unerringly to look for causes of outliers when they really exist and, hence, to decide whether previously given Alternative 1 is not the more plausible hypothesis to accept as compared to Alternative 2 in order that the most appropriate action in further data analysis may be taken. The procedures presented herein apply primarily to the simplest kind of experimental data, i.e., replicate measurements of some property of a given material or observations in a supposedly single random sample. Nevertheless, the tests suggested do cover a wide enough range of cases in practice to have rather broad utility.

When the skilled experimenter is clearly aware that a gross deviation from prescribed experimental procedure has taken place, the resultant observations should be discarded whether or not they agree with the rest of the data and without recourse to statistical tests for outliers. If a reliable correction procedure, for example, for temperature, is available, the observation may sometimes be corrected and retained.

In many cases evidence of deviation from prescribed procedure will consist primarily of the discordant value itself. In such cases it is advisable to adopt a cautious attitude. Use of one of the criteria discussed subsequently will sometimes permit a clear-cut decision to be made. In doubtful cases the experimenter's judgment will have considerable influence. When the experimenter cannot identify abnormal conditions, he should at least report the discordant values and indicate to what extent they have been used in the analysis of the data.

Thus for purposes of orientation relative to the overall problem of experimentation, our position on the matter of screening samples for outlying observations is precisely as follows:

1. Physical Reason Known or Discovered for Outlier(s):
    a. Reject observation(s).
    b. Correct observation(s) on physical grounds.
    c. Reject it (them) and possibly take additional observation(s).
2. Physical Reason Unknown—Use Statistical Test:
    a. Reject observation(s).
    b. Correct observation(s) statistically.
    c. Reject it (them) and possibly take additional observation(s).
    d. Employ truncated or censored sample theory not involving the suspected outliers for estimation purposes (Chapter 7).

The statistical test may always be used to lend support to a judgment that a physical reason does actually exist for an outlier, or the statistical criterion may be used routinely as a basis on which to initiate action to find a physical cause.

Before proceeding to the presentation and discussion of statistical significance tests for detecting outlying observations, we will cover a very important topic—namely, that of the mathematical bounds on certain of the key sample statistics. In other words, the statistical tests of significance will cover the cases in which we deal with or detect unusually large "random" variations, and there also actually exist some "mathematical limits" on the sample values or statistics themselves without any reference to random variations. These conditions will, in fact, have direct bearings on the suitability of the statistical tests of significance concerning whether they are even mathematically possible. For example, if for some given sample size there is an upper or mathematical bound on the deviation of the largest observation from the sample mean, there is no point in testing it statistically using the random sample theory to detect whether it is more deviant than that bound since this would be meaningless. We now discuss the mathematical bounds.

3-2 PRELIMINARIES AND MATHEMATICAL BOUNDS OF INTEREST

3-2.1 DESIGNATION OF THE SAMPLE

Ordinarily, in our procedures for detecting outlying observations in samples, we consider that a random sample of size n has been drawn from a population—almost always assumed to be a Gaussian or normal universe—and then a significance test will be carried out to judge whether or not, for example, the largest observation is too high or the smallest observation too low. However, for our discussion of mathematical bounds, we do not need to have any reference whatever to either a random sample or a normal universe.

We will designate the sample in the order the observations were drawn by

$$x_1', x_2', x_3', \ldots, x_i', \ldots, x_n'.$$

However, since we will be concerned almost exclusively with ordered sample observations, the sample values are listed as

$$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_i \leq \cdots \leq x_n$$

where

x_n = largest observation in the sample.
x_1 = smallest observation in the sample.

The sample mean $\bar{x}$ is given by

$$\bar{x} = \sum_{i=1}^{n} x_i/n = \Sigma x_i/n, \tag{3-1}$$

and the sample variance s^2 based on $(n-1)$ degrees of freedom (df) is given by

$$s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1) = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)} = \frac{A_{xx}}{n(n-1)} \tag{3-2a}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2/[2n(n-1)]. \tag{3-2b}$$

Eq. 3-2b for s^2 is especially of interest. Because the observations x_i and x_j $(i \neq j)$ are independent, it is easier to take expected values of that particular form, and if one of the observations, say x_k, $k \neq i$, is an outlier, the absolute difference $|x_i - x_k|$ would be large in comparison to other absolute differences not involving x_k.

Finally, we will make use of the maximum dispersion or sample range w given by

$$w = x_n - x_1, \tag{3-3}$$

i.e., the largest minus the smallest observations.

With these definitions, we may now give several mathematical bounds of interest.

3-2.2 BOUNDS FOR THE RATIO OF THE SAMPLE RANGE TO THE SAMPLE STANDARD DEVIATION

G. W. Thomson (Ref. 1) has determined the upper and lower mathematical bounds of the ratio w/s of the sample range to the standard deviation. We quote from his paper (Ref. 1):

"It can readily be shown that the upper and lower bounds of w/s for samples from any population with nonzero variance arise from certain simple configurations of the sample points. The upper bound, which corresponds to minimum s for a given range w, results from the arrangement with $(n-2)$ of the points at the sample mean and the other two points at equal distances from the mean. The lower bound, which corresponds to maximum s for a given w, results from the concentration of half of the sample points at one extreme and the other half (plus one, if the sample size is odd) of the sample points at the other extreme. The numerical values of the bounds can be shown to be: . . .

$$\left.\begin{array}{l} 2\sqrt{(n-1)/n}\text{ , for } n \text{ even} \\ 2\sqrt{n/(n+1)}\text{ , for } n \text{ odd} \end{array}\right\} \leq w/s \leq \sqrt{2(n-1)}\text{ . "} \tag{3-4}$$

We will illustrate the inequality (Eq. 3-4) with an example.

Example 3-1:

From the data of Table 2-2 we found that instrument I_1 had the largest standard deviation in errors of measurement. Hence, we calculate the ratio of the range to standard deviation and check with the bounds of Eq. 3-4 to see whether there is possibly an error in computation.

We see from Eq. 2-9 that

$$S_r^2 = 0.04714, \text{ and therefore, our } s = S_r = 0.2171.$$

Furthermore, from either Table 2-1 or Table 2-2 we note the largest reading for instrument I_1 is 10.32, and the smallest reading is 9.44, or the sample range is $w = 0.88$. Hence the quantity $w/s = 0.88/0.2171 = 4.053$, whereas the upper bound is

$$\sqrt{2(n-1)} = 7.62$$

and the lower bound is

$$2\sqrt{(n-1)/n} = 1.97$$

so that neither bound is reached, and "everything is go" to test for statistical outliers!

Since the standard deviation is the most efficient estimate of dispersion, but is more difficult to calculate, statisticians have often determined the range and used the bounds of Eq. 3-4 as a numerical check for wild values of the sample standard deviation.

3-2.3 BOUNDS FOR THE RESIDUALS OR DEVIATIONS FROM THE SAMPLE MEAN

In a 1968 paper titled "How Deviant Can You Be?", Nobel Prize winner Paul A. Samuelson (Ref. 2) studied maximum deviations from the sample and population means and showed that for a finite universe of N items, no value can lie more than $\sqrt{(N-1)}$ standard deviations away from the mean. Samuelson also showed for the sample standard deviation s' based on the number of sample items n, instead of $(n-1)$ df, that

$$\max|x_i - \bar{x}| \leq \sqrt{n-1}\, s' \tag{3-5}$$

where the sample standard deviation s' based on a total sample size n is

$$s' = \sqrt{\Sigma(x_i - \bar{x})^2/n}. \tag{3-6}$$

The conversion of s, from Eq. 3-2, to s' is given by

$$s = \sqrt{n/(n-1)}\, s' \tag{3-7}$$

and hence in terms of s, we also have that

$$\max|x_i - \bar{x}| \leq [(n-1)/\sqrt{n}]\, s. \tag{3-8}$$

Samuelson (Ref. 2) furthermore points out that the inequality (Eq. 3-5) may be sharpened in only special cases or restrictions:

"Thus, if the probability distribution is known to be symmetric, the greatest relevant deviant will be found where all but two of the observations are clustered halfway between the remaining two, and for a symmetric distribution the above theorem [our inequality (Eq. 3-5) using s'] can have $\sqrt{N-1}$ replaced by $\sqrt{N/2}$, a definite improvement when $N > 2$."

It is well-known that for any population with mean μ and standard deviation σ, the Tchebycheff inequality (TI) states that

$$Pr[|(x-\mu)/\sigma| \geq L] \leq 1/L^2 \tag{3-9}$$

where

L = selected "limit".

Samuelson (Ref. 2) applies this to a finite universe of N items by equating $1/L^2$ to $1/N$ to give

$$Pr[|(x-\mu)/\sigma| \geq \sqrt{N}] \leq 1/N \tag{3-10}$$

which, for example, states that for a universe of only 2 items not more than one of the observations can lie more than 1.414 standard deviations away from the mean with the probability 0.5. The inequality (Eq. 3-5) is much sharper, however, because it says that no observation may lie more than just 1.00 standard deviation from the mean.

Samuelson (Ref. 2) summarizes his results in terms of the following two theorems and a final summary:

"*Theorem.* Of N observations, no r (of them) [r = number less than N] can be more than the following number of standard deviations from the mean:

$\sqrt{N/r}$ for r an even number,

and

$(N-1)/\sqrt{(Nr-1)}$ for r an odd number.

"*Theorem*: No one of the N observations can be more than N mean absolute deviations away from the median.

"*Final Summary*: Although Tchebycheff's inequality cannot, in general, be improved upon, for universes (or samples) known to consist of a finite number of items N, an improvement on Tchebycheff's inequality is possible when dealing with r of N items, r being odd, but with the relative amount of improvement $\rightarrow 0$ as $N \rightarrow \infty$."

In a fundamental and very important paper, which appeared in 1936, Pearson and Chandra Sekar (Ref. 3) studied the recommendation of W. R. Thompson (Ref. 4) for detecting outliers in a sample based on the use of an arbitrary x_i selected at random from a sample of size n and the criterion

$$t_i = (x_i - \bar{x})/s'. \tag{3-11}$$

In particular, Pearson and Chandra Sekar (Ref. 3) were interested in the possible use of Eq. 3-11 and its efficiency in testing for outliers in the presence of more than a single outlier. They found, for example, that if the significance level of 0.10 (10%) were used, involving the risk of rejecting one observation in every 10 samples when the null hypothesis H_0 is true, then under no circumstances could one reject more than one observation until a sample of size $n = 11$ is reached, and one cannot reject more than two observations until $n = 22$ is reached; no more than three observations until $n = 33$, etc. This led Pearson and Chandra Sekar to make a thorough study of the mathematical bounds on the sample values since the statistical frequencies of acceptance and rejection from random sample theory may be spuriously interpreted.

Pearson and Chandra Sekar (Ref. 3) considered the n values of the t_i in a sample arranged in descending order of absolute magnitude as

$$|t_1| \geq |t_2| \geq \cdots \geq |t_n| \tag{3-12}$$

and also the n values of the t_i arranged in magnitude considering sign as

$$t^{(1)} \geq t^{(2)} \geq \cdots \geq t^{(n)}. \tag{3-13}$$

In an appendix to Ref. 3, J. M. C. Scott presented the following information concerning bounds that may be of some possible interest:

$$\max|t_i| = \sqrt{n(n-i)/[i(n-i)+1]}, \text{ if } i \text{ odd and } i < n \tag{3-14}$$

(or $\max|t_1| = \sqrt{n-1}$ as Samuelson (Ref. 2) later showed)

$$\max|t_n| = \sqrt{(n-1)/(n+1)}, \text{ if } i = n \text{ and } i \text{ is odd} \tag{3-15}$$

$$\max|t_i| = \sqrt{n/i}, \text{ if } i \text{ is even.} \tag{3-16}$$

The quantities $t^{(2)}$ and $t^{(n-1)}$ also reach into the tails of distributions of interest, as J. M. C. Scott (Appendix, Ref. 3) shows that

$$\max t^{(2)} = \sqrt{(n-2)/2} \tag{3-17}$$

and

$$\min t^{(n-1)} = -\sqrt{(n-2)/2}. \tag{3-18}$$

We quote from Scott (Ref. 3):
"The maximum value of $|t_1|$ occurs when $(n-1)$ of the observations have the same (identical) value and the remaining observation any different value. The maximum $t^{(2)}$ occurs when $(n-2)$ observations have the same (identical) value and the other two have a different but common value, that is, $t^{(1)} = t^{(2)}$. The maximum $|t_2|$ occurs when $(n-2)$ observations have the same or identical value and the other two differ with $t_1 = -t_2$. The maximum $|t_3|$ occurs when $(n-3)$ observations have the same value and the other three differ with $t_1 = t_2 = -t_3$, etc." (This process continues similarly as described in Ref. 3, Appendix.)

Thus we see that Pearson and Chandra Sekar (Ref. 3), in fact, made a very substantial contribution to the problem of testing random sample values for outliers, especially for small sample size n. Indeed, the mathematical bounds will be the controlling conditions in some cases, and we should be aware of their effect, especially insofar as such bounds have rigid controls on random sampling distributions for testing outliers.

With these preliminaries on mathematical bounds, we will consider the sampling or probability distributions for the special cases of samples of size either $n = 2$ or $n = 3$.

3-3 SOME RELATIONSHIPS AND SAMPLING DISTRIBUTIONS FOR SAMPLES OF SIZE TWO OR THREE

3-3.1 RELATION BETWEEN THE RANGE AND STANDARD DEVIATION FOR A SAMPLE OF SIZE TWO

When $n = 2$, there is a special relation between the sample range and the two sample standard deviations, i.e.,

$$w = 2s' = \sqrt{2}\,s \qquad (n = 2 \text{ only}). \tag{3-19}$$

The relationship given by Eq. 3-19 is often of some practical interest. In fact, since the range and the two sample standard deviations differ only by constant factors, it is easy to establish the probability distribution of all three quantities. In this connection, it is well-known from statistical theory that, for any sample size and the assumption of sampling a normal population, the quantities

$$(n-1)s^2/\sigma^2 = ns'^2/\sigma^2 = \Sigma(x_i - \bar{x})^2/\sigma^2 = \chi^2(n-1). \tag{3-20}$$

Or, the total sum of squares (SS) about the sample mean divided by the population variance follows the chi-square distribution with $(n-1)$ df.

From Eq. 3-20 it is easily noted that when we have a sample of size $n = 2$,

$$s^2/\sigma^2 = 2s'^2/\sigma^2 = w^2/(2\sigma^2) = \chi^2(1) \tag{3-21}$$

or all of the first three quantities in Eq. 3-21 are distributed as chi-square with a single degree of freedom.

Moreover, from Eq. 3-21 it is easily seen that

$$s/\sigma = \sqrt{2}\, s'/\sigma = w/\sqrt{2}\,\sigma = \chi(1), \tag{3-22}$$

or the square roots of the first three quantities are distributed as chi with 1 df.

This means that

$$Pr[s/\sigma < \chi_0] = Pr[s'/\sigma < \chi_0/\sqrt{2}\,] = Pr[w/\sigma \leq \sqrt{2}\,\chi_0]$$
$$= 2\int_{-\infty}^{\chi_0} (1/\sqrt{2\pi})\exp(-t^2/2)dt - 1,\ \chi_0 \geq 0 \tag{3-23}$$

which is in terms of the standardized normal integral

3-3.2 THE RANGE FOR SAMPLES OF SIZE THREE AND PROPERTIES OF THE TWO CLOSEST OF THREE OBSERVATIONS

The case of a sample of size three ($n = 3$) from a normal population is also of some special practical interest concerning the problem of outliers. To begin with, the ratio of the sample range to the sample standard deviation takes on a rather simple distributional form, and historically, there has been much interest in samples of size three from the standpoint of checking results. Thus many experimenters, especially chemists, have reasoned as follows: "If I take only one observation, then I can't be sure it is a good value. If I take two observations, then I can't know which one is correct either. But if I take three observations, then I can always select the closest two of the three and depend on them!"

The range w_3 of a sample of three observations is

$$w_3 = x_3 - x_1 \tag{3-24}$$

i.e., the largest minus the smallest of the observations.

It can be shown (see for example Ref. 5, p. vii, Eq. 12, and p. xxxiii, Eq. 46, that the probability distribution of w_3/σ can be related directly to the bivariate normal distribution. In fact, for samples of size $n = 3$

$$Pr[w_3/\sigma \leq w_0] = 12V(w_0/\sqrt{2},\ w_0/\sqrt{6}\,) \tag{3-25}$$
$$= 12\int_0^{w_0/\sqrt{2}} \int_0^{x/\sqrt{3}} (1/2\pi)\exp[-(x^2 + y^2)/2]dxdy$$

and it may be determined directly from Table III of Ref. 5.

The probability integral of the range for sample sizes of $n = 2$ (1)20, including $n = 3$, has been tabulated by Pearson and Hartley in Ref. 6.

As a result of intense interest on the part of scientific and engineering personnel, especially chemists, Lieblein (Ref. 7) carried out an excellent study on the properties of certain sample statistics involving the closest pair of observations in a sample of size three. This is especially important since there is clearly a very natural tendency to quote, use, and depend on only the closest two of three observations and to brand the remaining one as being discrepant, or an "outlier". Lieblein describes the condition quite aptly in the abstract or summary of his paper (Ref. 7) as follows:

"Triplicate readings are of wide occurrence in experimental work. Occasionally, however, only the closest pair of a triad is used, and the outlying high or low one discarded as evidencing some gross error. The present paper presents a mathematical investigation leading to precise determination of some of the biases that result from such selection. This project was suggested by certain experiments involving random sampling numbers and analysis of published chemical determinations. The theoretical findings agree closely with the empirical results and imply that selected pairs not only tend to overestimate considerably the precision of the experimental procedure, but also result in less accurate determinations.".

Lieblein's paper (Ref. 7) is highly recommended for study by experimental investigators in all fields of application since the investigators may be often throwing away important information in the sample and, hence, possibly render bias to their conclusions. For our purposes, however, we will limit our coverage to the sampling distributions of normal samples of size three for (1) the ratio of the difference between the closest two of three observations and the sample range and (2) the ratio of the sample range to the sample standard deviation. Thus the three ordered observations are

$$x_1 \leq x_2 \leq x_3$$

and, as Lieblein did, we designate these three (not ordered) values as

$$x', x'', x'''$$

where x' and x'' are the *closest* two of the three, and we take $x' \geq x''$ for convenience. Lieblein then finds the probability distribution function (pdf) of

$$y = (x' - x'') / (x_3 - x_1) \tag{3-26}$$

for sample of $n = 3$ from a normal parent to be simply

$$f(y) = 3\sqrt{3}/[\pi(y^2 - y + 1)],\ 0 \leq y \leq 1/2. \tag{3-27}$$

We note that the sample statistic y in Eq. 3-26 does not depend on any nuisance population parameters and is completely independent of origin and scale effects. Thus for random samples of three from an assumed normal population, Eq. 3-26 may be calculated to discern whether the closest two observations are actually too close or too far apart by referring the calculated value to a table of percentage points.

The cumulative distribution of y in Eq. 3-26 is (Ref. 7)

$$Pr[y \leq y_0] = F(y_0) = (6/\pi) \text{ arc tan } [(2y_0 - 1)/\sqrt{3}\,] + 1^* \tag{3-28}$$

where

y_o = any upper limit.

The mean $E(y)$ and standard deviation $\sigma(y)$ of y are

$$E(y) = 0.2621 \tag{3-29}$$

and

$$\sigma(y) = 0.1428. \tag{3-30}$$

The lower 1% probability level of Eq. 3-28 is $y_0 = 0.00603$, and the lower 5% level is at $y_0 = 0.02979$ (Ref. 7) for judging whether the two closest observations are "unusually close", so that the third one is an "outlier". If y of Eq. 3-26 does not fall below one of these selected values, the remaining observation should not be suspected.

For samples of size three, a paper by Anscombe and Barron (Ref. 8) is also of particular interest because it discusses the choice of an outlier rejection criterion in terms of the effect of it on the mean square error of estimates of population parameters, e.g., the mean.

Finally, for samples of three observations the distribution of the sample range divided by the sample standard deviation, i.e., w/s, may be of particular interest and, in fact, takes on a rather simple form. Thomson (Ref. 1) points out in this connection that the upper α_{1-F} percentage point of w/s is determined simply from

$$\alpha_{1-F} = 2\cos[30°(1 - F)] \tag{3-31}$$

*The arc tan is in radians. When arc tan is expressed in deg, the constant $6/\pi$ must be changed to $1/30$.

where

F = cumulative relative frequency.

Thus if we want the upper 5% point or the 95% cumulative level, we set $F = 0.95$ and find that

$$\alpha_{0.95} = \text{Upper } \alpha_{0.05} = 1.9993.$$

Lower percentage points are obtained by putting $F < 0.50$, e.g., $F = 0.05$ in Eq. 3-31 gives the lower 5% level—or actually the 5% level—as

$$\text{Lower } \alpha_{0.05} = 2\cos[30°(0.95)] = 1.75763.$$

Unfortunately, if w/s is significantly low (or high), it would not reveal whether x_1 or x_3 is an outlier. Thus Lieblein's closest two out of three test, or Eq. 3-26, would be best for this. See Example 3-2 for an illustration of Lieblein's procedure.

Example 3-2:

To illustrate Lieblein's "closest pair of three" statistical test, let us take the data on the fourth round of Table 2-2. In this particular case the measured times for observers I_1, I_2, and I_3 are 9.79, 9.71, and 9.70 s, respectively. Is there any evidence that I_1's reading of 9.79 is an outlier?

We note in this connection that

$$9.70 < 9.71 < 9.79,$$

so that the range $w = x_3 - x_1 = 9.79 - 9.70 = 0.09$. Also $x' = 9.71$ and $x'' = 9.70$, so that $x' - x'' = 0.01$. Thus from Eq. 3-26 we see that Lieblein's

$$y = \frac{9.71 - 9.70}{9.79 - 9.70} = 0.01/0.09 = 0.111$$

and from Eq. 3-28

$$Pr[y \leq 0.111] = 0.19$$

which does not fall in the range of a significant probability, i.e., $Pr \leq 0.05$, for example. Therefore, we conclude that the closest two values, 9.70 and 9.71, are not so close as to indicate that 9.79 should be discarded. Also this example points out that, as Lieblein has indicated, if only the closest two values of the three were used, we would be discarding too often an apparently good observation due to random sampling.

3-4 BASIS OF STATISTICAL CRITERIA FOR OUTLIERS

We will now develop sample criteria for testing the significance of the outlying or remote values for general sample sizes—i.e., not only for $n = 2$ or 3 as previously stated, but also for any greater sample size as well. In fact, the coverage that follows represents the more usual cases that will occur in practice.

There are a number of criteria for testing outliers. In all of these the doubtful observation is included in the calculation of the numerical value of a sample criterion (or statistic). The numerical value is then compared with a critical value based on the theory of random sampling to determine whether the doubtful observation is to be retained or rejected. The critical value is that value of the sample criterion that would be exceeded by chance with some specified (small) probability on the assumption that all the observations did indeed constitute a random sample from a common system of causes, a single parent population, distribution, or universe. The specified small probability is called the significance level or percentage point

and can be thought of as the risk of erroneously rejecting a good observation. It becomes clear, therefore, that if there exists a real shift or change in the value of an observation that arises from nonrandom causes—human error, loss of calibration of instrument, change of measuring instrument, or even change of time of measurements, etc.—the numerical value of the sample criterion used would exceed the critical value based on random sampling theory. Tables of critical values are usually given for several different significance levels, for example, 5% or 1%. For statistical tests of outlying observations, it is generally recommended that a low significance level, such as 1%, be used and that significance levels greater than 5% would not be common practice. In this chapter we will usually illustrate the use of the 5% significance level. Proper choice of a significance level depends on the particular problem, just what may be involved, and the risk that one is willing to take in rejecting a good observation—i.e., whether the null hypothesis stating "all observations in the sample come from the same normal population" may be properly assumed.

It should be pointed out that almost all criteria for outliers are based on an assumed underlying normal (Gaussian) population, universe, or distribution. When the data are not normally or approximately normally distributed, the probabilities associated with these tests will be different. Until such time as criteria not sensitive to the normality assumption are developed, the experimenter should be cautioned against interpreting the probabilities too literally.

Although our primary interest is to detect outlying observations, we remark that some of the statistical criteria presented may also be used to test the hypothesis of normality or that the random sample taken did indeed come from a normal, or Gaussian, population. For all practical purposes the end result is the same, i.e., we really wish to know whether we ought to proceed as if we have a sample of homogeneous observations—i.e., no outlying observations—from the same (normal) universe.

3-5 RECOMMENDED OUTLIER DETECTION CRITERIA FOR SINGLE SAMPLES

3-5.1 TESTS FOR EITHER THE HIGHEST OR LOWEST OBSERVATION

Let the sample of n observations be denoted in order of increasing magnitude $x_1 \leq x_2 \leq x_3 \leq \leq x_n$. The x_1 or x_n denotes the doubtful value, i.e., the smallest or largest value. The test criterion for the largest item T_n, recommended for testing whether or not the largest observation is an outlier, based on the work of Grubbs (Refs. 9, 10, 11, and 12), is as follows:

$$T_n = \frac{x_n - \bar{x}}{s} \tag{3-32}$$

where

$\bar{x}$ = arithmetic average of all n values

s = estimate of the population standard deviation based on the sample data calculated as follows:

$$s = \left[\frac{\Sigma(x_i - \bar{x})^2}{n-1}\right]^{1/2} = \left[\frac{n\Sigma x_i^2 - (\Sigma x_i)^2}{n(n-1)}\right]^{1/2} \tag{3-33}$$

If x_1, the smallest value, rather than x_n, is the doubtful value, the test criterion (Refs. 9, 10, 11, and 12) is

$$T_1 = \frac{\bar{x} - x_1}{s} \quad . \tag{3-34}$$

The critical values for either case, for the 1 and 5% levels of significance, from Grubbs and Beck (Ref. 13), are given in Table 3-1. Table 3-1 gives the one-sided significance levels. In many previous treatments of outliers, the tables listed values of significance levels double those in the accompanying tables since it was considered that the experimenter would test either the lowest or highest observation (or both) for statistical significance. However, to be consistent with actual practice and in an attempt to avoid any further

TABLE 3-1

CRITICAL VALUES FOR T (ONE-SIDED TEST OF T_1 OR T_n) WHEN THE STANDARD DEVIATION IS CALCULATED FROM THE SAME SAMPLE (Ref. 13)

No. Obs. n	Upper 0.1% Sig. Level	Upper 0.5% Sig. Level	Upper 1% Sig. Level	Upper 2.5% Sig. Level	Upper 5% Sig. Level	Upper 10% Sig. Level
3	1.155	1.155	1.155	1.155	1.153	1.148
4	1.499	1.496	1.492	1.481	1.463	1.425
5	1.780	1.764	1.749	1.715	1.672	1.602
6	2.011	1.973	1.944	1.887	1.822	1.729
7	2.201	2.139	2.097	2.020	1.938	1.828
8	2.358	2.274	2.221	2.126	2.032	1.909
9	2.492	2.387	2.323	2.215	2.110	1.977
10	2.606	2.482	2.410	2.290	2.176	2.036
11	2.705	2.564	2.485	2.355	2.234	2.088
12	2.791	2.636	2.550	2.412	2.285	2.134
13	2.867	2.699	2.607	2.462	2.331	2.175
14	2.935	2.755	2.659	2.507	2.371	2.213
15	2.997	2.806	2.705	2.549	2.409	2.247
16	3.052	2.852	2.747	2.585	2.443	2.279
17	3.103	2.894	2.785	2.620	2.475	2.309
18	3.149	2.932	2.821	2.651	2.504	2.335
19	3.191	2.968	2.854	2.681	2.532	2.361
20	3.230	3.001	2.884	2.709	2.557	2.385
21	3.266	3.031	2.912	2.733	2.580	2.408
22	3.300	3.060	2.939	2.758	2.603	2.429
23	3.332	3.087	2.963	2.781	2.624	2.448
24	3.362	3.112	2.987	2.802	2.644	2.467
25	3.389	3.135	3.009	2.822	2.663	2.486
26	3.415	3.157	3.029	2.841	2.681	2.502
27	3.440	3.178	3.049	2.859	2.698	2.519
28	3.464	3.199	3.068	2.876	2.714	2.534
29	3.486	3.218	3.085	2.893	2.730	2.549
30	3.507	3.236	3.103	2.908	2.745	2.563
31	3.528	3.253	3.119	2.924	2.759	2.577
32	3.546	3.270	3.135	2.938	2.773	2.591
33	3.565	3.286	3.150	2.952	2.786	2.604
34	3.582	3.301	3.164	2.965	2.799	2.616
35	3.599	3.316	3.178	2.979	2.811	2.628
36	3.616	3.330	3.191	2.991	2.823	2.639
37	3.631	3.343	3.204	3.003	2.835	2.650
38	3.646	3.356	3.216	3.014	2.846	2.661
39	3.660	3.369	3.228	3.025	2.857	2.671
40	3.673	3.381	3.240	3.036	2.866	2.682
41	3.687	3.393	3.251	3.046	2.877	2.692
42	3.700	3.404	3.261	3.057	2.887	2.700
43	3.712	3.415	3.271	3.067	2.896	2.710
44	3.724	3.425	3.282	3.075	2.905	2.719
45	3.736	3.435	3.292	3.085	2.914	2.727
46	3.747	3.445	3.302	3.094	2.923	2.736
47	3.757	3.455	3.310	3.103	2.931	2.744
48	3.768	3.464	3.319	3.111	2.940	2.753

(cont'd on next page)

TABLE 3-1 (cont'd)

No. Obs. *n*	Upper 0.1% Sig. Level	Upper 0.5% Sig. Level	Upper 1% Sig. Level	Upper 2.5% Sig. Level	Upper 5% Sig. Level	Upper 10% Sig. Level
49	3.779	3.474	3.329	3.120	2.948	2.760
50	3.789	3.483	3.336	3.128	2.956	2.768
51	3.798	3.491	3.345	3.136	2.964	2.775
52	3.808	3.500	3.353	3.143	2.971	2.783
53	3.816	3.507	3.361	3.151	2.978	2.790
54	3.825	3.516	3.368	3.158	2.986	2.798
55	3.834	3.524	3.376	3.166	2.992	2.804
56	3.842	3.531	3.383	3.172	3.000	2.811
57	3.851	3.539	3.391	3.180	3.006	2.818
58	3.858	3.546	3.397	3.186	3.013	2.824
59	3.867	3.553	3.405	3.193	3.019	2.831
60	3.874	3.560	3.411	3.199	3.025	2.837
61	3.882	3.566	3.418	3.205	3.032	2.842
62	3.889	3.573	3.424	3.212	3.037	2.849
63	3.896	3.579	3.430	3.218	3.044	2.854
64	3.903	3.586	3.437	3.224	3.049	2.860
65	3.910	3.592	3.442	3.230	3.055	2.866
66	3.917	3.598	3.449	3.235	3.061	2.871
67	3.923	3.605	3.454	3.241	3.066	2.877
68	3.930	3.610	3.460	3.246	3.071	2.883
69	3.936	3.617	3.466	3.252	3.076	2.888
70	3.942	3.622	3.471	3.257	3.082	2.893
71	3.948	3.627	3.476	3.262	3.087	2.897
72	3.954	3.633	3.482	3.267	3.092	2.903
73	3.960	3.638	3.487	3.272	3.098	2.908
74	3.965	3.643	3.492	3.278	3.102	2.912
75	3.971	3.648	3.496	3.282	3.107	2.917
76	3.977	3.654	3.502	3.287	3.111	2.922
77	3.982	3.658	3.507	3.291	3.117	2.927
78	3.987	3.663	3.511	3.297	3.121	2.931
79	3.992	3.669	3.516	3.301	3.125	2.935
80	3.998	3.673	3.521	3.305	3.130	2.940
81	4.002	3.677	3.525	3.309	3.134	2.945
82	4.007	3.682	3.529	3.315	3.139	2.949
83	4.012	3.687	3.534	3.319	3.143	2.953
84	4.017	3.691	3.539	3.323	3.147	2.957
85	4.021	3.695	3.543	3.327	3.151	2.961
86	4.026	3.699	3.547	3.331	3.155	2.966
87	4.031	3.704	3.551	3.335	3.160	2.970
88	4.035	3.708	3.555	3.339	3.163	2.973
89	4.039	3.712	3.559	3.343	3.167	2.977
90	4.044	3.716	3.563	3.347	3.171	2.981
91	4.049	3.720	3.567	3.350	3.174	2.984
92	4.053	3.725	3.570	3.355	3.179	2.989
93	4.057	3.728	3.575	3.358	3.182	2.993
94	4.060	3.732	3.579	3.362	3.186	2.996
95	4.064	3.736	3.582	3.365	3.189	3.000
96	4.069	3.739	3.586	3.369	3.193	3.003
97	4.073	3.744	3.589	3.372	3.196	3.006
98	4.076	3.747	3.593	3.377	3.201	3.011

(cont'd on next page)

TABLE 3-1 (cont'd)

No. Obs. n	Upper 0.1% Sig. Level	Upper 0.5% Sig. Level	Upper 1% Sig. Level	Upper 2.5% Sig. Level	Upper 5% Sig. Level	Upper 10% Sig. Level
99	4.080	3.750	3.597	3.380	3.204	3.014
100	4.084	3.754	3.600	3.383	3.207	3.017
101	4.088	3.757	3.603	3.386	3.210	3.021
102	4.092	3.760	3.607	3.390	3.214	3.024
103	4.095	3.765	3.610	3.393	3.217	3.027
104	4.098	3.768	3.614	3.397	3.220	3.030
105	4.102	3.771	3.617	3.400	3.224	3.033
106	4.105	3.774	3.620	3.403	3.227	3.037
107	4.109	3.777	3.623	3.406	3.230	3.040
108	4.112	3.780	3.626	3.409	3.233	3.043
109	4.116	3.784	3.629	3.412	3.236	3.046
110	4.119	3.787	3.632	3.415	3.239	3.049
111	4.122	3.790	3.636	3.418	3.242	3.052
112	4.125	3.793	3.639	3.422	3.245	3.055
113	4.129	3.796	3.642	3.424	3.248	3.058
114	4.132	3.799	3.645	3.427	3.251	3.061
115	4.135	3.802	3.647	3.430	3.254	3.064
116	4.138	3.805	3.650	3.433	3.257	3.067
117	4.141	3.808	3.653	3.435	3.259	3.070
118	4.144	3.811	3.656	3.438	3.262	3.073
119	4.146	3.814	3.659	3.441	3.265	3.075
120	4.150	3.817	3.662	3.444	3.267	3.078
121	4.153	3.819	3.665	3.447	3.270	3.081
122	4.156	3.822	3.667	3.450	3.274	3.083
123	4.159	3.824	3.670	3.452	3.276	3.086
124	4.161	3.827	3.672	3.455	3.279	3.089
125	4.164	3.831	3.675	3.457	3.281	3.092
126	4.166	3.833	3.677	3.460	3.284	3.095
127	4.169	3.836	3.680	3.462	3.286	3.097
128	4.173	3.838	3.683	3.465	3.289	3.100
129	4.175	3.840	3.686	3.467	3.291	3.102
130	4.178	3.843	3.688	3.470	3.294	3.104
131	4.180	3.845	3.690	3.473	3.296	3.107
132	4.183	3.848	3.693	3.475	3.298	3.109
133	4.185	3.850	3.695	3.478	3.302	3.112
134	4.188	3.853	3.697	3.480	3.304	3.114
135	4.190	3.856	3.700	3.482	3.306	3.116
136	4.193	3.858	3.702	3.484	3.309	3.119
137	4.196	3.860	3.704	3.487	3.311	3.122
138	4.198	3.863	3.707	3.489	3.313	3.124
139	4.200	3.865	3.710	3.491	3.315	3.126
140	4.203	3.867	3.712	3.493	3.318	3.129
141	4.205	3.869	3.714	3.497	3.320	3.131
142	4.207	3.871	3.716	3.499	3.322	3.133
143	4.209	3.874	3.719	3.501	3.324	3.135
144	4.212	3.876	3.721	3.503	3.326	3.138
145	4.214	3.879	3.723	3.505	3.328	3.140
146	4.216	3.881	3.725	3.507	3.331	3.142
147	4.219	3.883	3.727	3.509	3.334	3.144

Reprinted with permission. Copyright © by the American Statistical Association.

misunderstanding, single-sided significance levels are tabulated herein so that both viewpoints can be represented. The user can then make his own judgments in his many individual applications.

The hypothesis that we are testing in every case is that all observations in the sample come from the same normal population. Let us adopt, for example, a significance level of 0.05 (or 0.01). If we are interested *only* in outliers that occur on the *high side*, we should always use the statistic $T_n = (x_n - \bar{x})/s$ (Eq. 3-32) and take as critical value the 0.05 (or 0.01) point of Table 3-1. On the other hand, if we are interested *only* in outliers occurring on the *low side*, we should always use the statistic $T_1 = (\bar{x} - x_1)/s$ (Eq. 3-34) and again take as a critical value the 0.05 (or 0.01) point of Table 3-1. Suppose, however, that we are interested in outliers occurring on *either side* but do not believe that outliers can occur on both sides simultaneously. We might believe that at some time during the experiment something possibly happened to cause an extraneous variation on the high side or on the low side but that it was very unlikely that two or more such events could have occurred: one being an extraneous variation on the high side *and* the other an extraneous variation on the low side. With this point of view we should use the statistic $T_n = (x_n - \bar{x})/s$ *or* the statistic $T_1 = (\bar{x} - x_1)/s$, whichever is larger. If in this instance we use the 0.05 point of Table 3-1 as our critical value, the true significance level would be twice 0.05 or 0.10. If we wish a significance level of 0.05 and not 0.10, we must, in this case, use as a critical value the 0.025 point of Table 3-1. Similar considerations apply to the other tests given in the sequel.

Example 3-3:

As an illustration of the use of T_n and Table 3-1, consider the following 10 observations on breaking strength (in pounds) of 0.104-in. hard-drawn copper wire arranged in increasing order: 568, 570, 570, 570, 572, 572, 572, 578, 584, 596. The doubtful observation is the high value, $x_{10} = 596$. Is the value of 596 significantly high?

The mean is $\bar{x} = 575.2$, and the estimated standard deviation is $s = 8.70$. We compute

$$T_{10} = \frac{596 - 575.2}{8.70} = 2.39.$$

From Table 3-1 for $n = 10$, note that a T_{10} as large as 2.39 would occur by chance with probability less than 0.05. In fact, so large a value would occur by chance not much more often than 1% of the time. Thus using the 5% level of significance, the weight of the evidence is against the doubtful value having come from the same population as the others (assuming the population is normally distributed). Investigation of the doubtful value on physical grounds is therefore indicated.

3-5.2 DIXON'S CRITERIA

An alternative system, the Dixon criteria (based entirely on ratios of differences between the observations), is described in the literature (Ref. 14). It may be used in cases where it is desirable to avoid calculation of the standard deviation s or where quick judgment is necessary. For the Dixon test the sample criterion, or statistic, changes with sample size. Table 3-2 gives the appropriate statistic to calculate and also gives the critical values of the statistic for the 1, 5, and 10% levels of significance.

Example 3-4:

As an illustration of the use of Dixon's test, consider again the observations on breaking strength given in Example 3-3, and suppose that a large number of such samples had to be screened quickly for outliers, and it was judged too time-consuming to compute s. Table 3-2 for $n = 10$ indicates use of

$$r_{11} = \frac{x_n - x_{n-1}}{x_n - x_2} \tag{3-35}$$

Thus for $n = 10$,

$$r_{11} = \frac{x_{10} - x_9}{x_{10} - x_2} \tag{3-36}$$

TABLE 3-2
DIXON CRITERIA FOR TESTING OF EXTREME OBSERVATION (SINGLE SAMPLE)[a] (Ref. 14)

n	Criterion	Significance Level 10%	5%	1%
3	$r_{10} = (x_2 - x_1)/(x_n - x_1)$ if smallest value is suspected;	0.886	0.941	0.988
4	$= (x_n - x_{n-1})/(x_n - x_1)$ if largest value is suspected.	0.679	0.765	0.889
5		0.557	0.642	0.780
6		0.482	0.560	0.698
7		0.434	0.507	0.637
8	$r_{11} = (x_2 - x_1)/(x_{n-1} - x_1)$ if smallest value is suspected;	0.479	0.554	0.683
9	$= (x_n - x_{n-1})/(x_n - x_2)$ if largest value is suspected.	0.441	0.512	0.635
10		0.409	0.477	0.597
11	$r_{21} = (x_3 - x_1)/(x_{n-1} - x_1)$ if smallest value is suspected;	0.517	0.576	0.679
12	$= (x_n - x_{n-2})/(x_n - x_2)$ if largest value is suspected.	0.490	0.546	0.642
13		0.467	0.521	0.615
14	$r_{22} = (x_3 - x_1)/(x_{n-2} - x_1)$ if smallest value is suspected;	0.492	0.546	0.641
15	$= (x_n - x_{n-2})/(x_n - x_3)$ if largest value is suspected.	0.472	0.525	0.616
16		0.454	0.507	0.595
17		0.438	0.490	0.577
18		0.424	0.475	0.561
19		0.412	0.462	0.547
20		0.401	0.450	0.535
21		0.391	0.440	0.524
22		0.382	0.430	0.514
23		0.374	0.421	0.505
24		0.367	0.413	0.497
25		0.360	0.406	0.489

[a] $x_1 \leq x_2 \leq \cdots \leq x_n$

Reprinted with permission. Copyright © by Biometrika Trustees.

For the measurements of breaking strength in this example,

$$r_{11} = \frac{596 - 584}{596 - 570} = 0.462,$$

which is a little less than 0.477, the 5% critical value for $n = 10$. Therefore, under the Dixon criterion, we should *not* consider this observation as an outlier at the 5% level of significance. These results illustrate how borderline cases may be accepted under one test but rejected under another.

It should be remembered, however, that the T statistic previously discussed is the best one to use for the single outlier case, and final statistical judgment should be based on it. See, for example, Ferguson (Refs. 15 and 16). (The advent of the modern, scientific pocket calculator may reduce the need for the "quick" Dixon ratios.)

Further examination of the sample observations on breaking strength of hand-drawn copper wire indicates that none of the other values need testing for rejection.

With experience we may just look at the sample values to observe whether an outlier is present. However, strictly speaking, the statistical test should be applied to all samples under examination to guarantee

the significance levels used. Comments are made later concerning multiple tests for outliers in a single sample since it changes the overall significance level.

A test equivalent to T_n(or T_1) based on the sample sum of squared deviations from the mean for all the observations and the sum of squared deviations omitting the outlier is given by Grubbs in Ref. 9.

3-5.3 OUTLIER TEST FOR SMALLEST AND LARGEST OBSERVATIONS

The next type of problem to consider is the case in which there is the possibility of two outlying observations, i.e., the least and the greatest observations in a sample. (The problem of testing the two highest or the two lowest observations is considered in par. 3-5.4.) To test the least and the greatest observations simultaneously as probable outliers in a sample, we use the ratio of the sample range to the sample standard deviation test of David, Hartley, and Pearson (Ref. 17). The significance levels for this sample criterion are given in Table 3-3. Alternatively, the largest residuals test of Tietjen and Moore (Ref. 18) could be used, as in par. 3-5.5.2. The procedure for the test of David, Hartley, and Pearson is explained by Example 3-5.

Example 3-5:

There is one rather famous set of observations that a number of writers on the subject of outlying observations have referred to in applying their various tests for outliers. This classic set consists of a sample of 15 observations of the vertical semidiameters of Venus made by Lieutenant Herndon in 1846 (Ref. 19). In the reduction of the observations, the following residuals were found, which have been arranged in ascending order of magnitude:

−1.40 in.	−0.24	−0.05	0.18	0.48
−0.44	−0.22	0.06	0.20	0.63
−0.30	−0.13	0.10	0.39	1.01.

The deviations −1.40 and 1.01 appear to be outliers. Here the suspected observations lie at each end of the sample. Much less work has been accomplished for the case of outliers at both ends of the sample than for the case of one or more outliers at only one end of the sample. This is not necessarily because the one-sided case occurs more frequently in practice but because two-sided tests are somewhat more difficult with which to deal. For a high and a low outlier in a single sample, we give two procedures. The first is a combination of tests, which includes the test of David, Hartley, and Pearson (Ref. 17). The second is a single test of Tietjen and Moore (Ref. 18), discussed in par. 3-5.5.2, which may have nearly optimum properties.

For the observations on the semidiameter of Venus previously stated, all the information on the available measurement errors is contained in the sample of 15 residuals. In cases like this in which no independent estimate of variance is available (i.e., we still have the single sample case), a useful statistic is the ratio of the range of the observations to the sample standard deviation (David, Hartley, and Pearson, Ref. 17):

$$\frac{w}{s} = \frac{x_n - x_1}{s}, \quad x_1 \leq x_2 \leq \cdots \leq x_n \tag{3-37}$$

where

s is as in Eq. 3-33.

If x_n were about as far above the mean $\bar{x}$ as x_1 is below $\bar{x}$ and if w/s were to exceed the chosen critical value from Table 3-3, one would conclude that *both* the doutful values could be outliers. If, however, x_1 and x_n were displaced from the mean by rather different amounts, then some further test would have to be made to decide whether to reject as outlying only the lowest value, only the highest value, or both the lowest and highest values.

For this example the mean of the residuals or deviations is $\bar{x} = 0.018$, the sample standard deviation $s = 0.551$, and the David, Hartley, and Pearson statistic (Ref. 17) is

TABLE 3-3

CRITICAL VALUES FOR w/s (RATIO OF RANGE TO SAMPLE STANDARD DEVIATION)[a] (Ref. 17)

Number of Observations n	5% Significance Level	1% Significance Level	0.5% Significance Level
3	2.00	2.00	2.00
4	2.43	2.44	2.45
5	2.75	2.80	2.81
6	3.01	3.10	3.12
7	3.22	3.34	3.37
8	3.40	3.54	3.58
9	3.55	3.72	3.77
10	3.68	3.88	3.94
11	3.80	4.01	4.08
12	3.91	4.13	4.21
13	4.00	4.24	4.32
14	4.09	4.34	4.43
15	4.17	4.43	4.53
16	4.24	4.51	4.62
17	4.31	4.59	4.69
18	4.38	4.66	4.77
19	4.43	4.73	4.84
20	4.49	4.79	4.91
30	4.89	5.25	5.39
40	5.15	5.54	5.69
50	5.35	5.77	5.91
60	5.50	5.93	6.09
80	5.73	6.18	6.35
100	5.90	6.36	6.54
150	6.18	6.64	6.84
200	6.38	6.85	7.03
500	6.94	7.42	7.60
1000	7.33	7.80	7.99

[a] $w = x_n - x_1,\ x_1 \leq x_2 \leq \cdots \leq x_n$

$s = \sqrt{\Sigma(x_i - \bar{x})^2/(n-1)}$

Reprinted with permission. Copyright © by Biometrika Trustees.

$$w/s = \frac{1.01 - (-1.40)}{0.551} = \frac{2.41}{0.551} = 4.374.$$

From Table 3-3 for $n = 15$, we see that the value of $w/s = 4.374$ falls between the critical values for the 1 and 5% levels. If the test were being run at the 5% level of significance, we would conclude that this sample contains one or more outliers. The lowest measurement, −1.40 in., is 1.418 below the sample mean; the highest measurement, 1.01 in., is 0.992 above the mean. Since, however, these extremes are not

symmetric about the mean, either both extremes are outliers, or only −1.40 is an outlier. That −1.40 is an outlier can be verified by use of the T_1 statistic of Eq. 3-34. We have from Eq. 3-34 that

$$T_1 = (\bar{x} - x_1)/s = \frac{0.018 - (-1.40)}{0.551} = 2.574.$$

This value is greater than the critical value of 2.409 from Table 3-1 for the 5% level; therefore, we should look for the cause of this or reject −1.40. Since we have decided that −1.40 is an outlier, we use the remaining 14 observations and test the upper extreme observation 1.01 either with the criterion (Eq. 3-32)

$$T_n = \frac{x_n - \bar{x}}{s}$$

or with Dixon's r_{22}. Omitting −1.40 and renumbering the observations, we compute

$$\bar{x} = \frac{1.67}{14} = 0.119, s = 0.401$$

and

$$T_{14} = \frac{1.01 - 0.119}{0.401} = 2.22.$$

From Table 3-1 for $n = 14$ we find that a value as large as 2.22 would occur by chance more than 5% of the time, so we should retain the value 1.01 in further calculations. For further information we calculate Dixon's

$$r_{22} = \frac{x_{14} - x_{12}}{x_{14} - x_3} = \frac{1.01 - 0.48}{1.01 + 0.24} = \frac{0.53}{1.25} = 0.424.$$

From Dixon's Table 3-2 for $n = 14$, we see that the 5% critical value for r_{22} is 0.546. Since our calculated value (0.424) is less than the critical value, we also retain 1.01 by Dixon's test, and no further values would be tested in this sample.

It should be noted that in a multiplicity of tests of this kind, the final, overall significance level will be somewhat less than that used in the individual tests since we are offering more than one chance of accepting the sample as one produced by a random operation.* It is not our purpose to cover the theory of multiple tests very extensively because it introduces a broad subject area although we will give some coverage of multiple-type tests as required in pars. 3-5.5.2 and 3-5.5.3.

Finally, we should remark at this point that we have begun to reject some of the suspected outliers in our examples. To many experimental investigators, the matter of just rejecting observations on statistical grounds and depending on inferences from the remaining "statistically homogeneous" values "sounds a very sour note" indeed. We agree that we must be very careful about rejecting observations, including perhaps the outlying ones, unless we can very definitely establish that they are due to errors of measurement, for example, and do not represent the true characteristics of the physical process we are sampling or investigating. Actually, data are taken, hopefully, to make further inferences from our investigations or to place our findings in a generalized framework. Thus we desire to estimate population means, standard deviations, and other characteristics of the universe we are sampling, and the rejection of observations will very definitely have an important effect on any such inferences. For this reason, we will discuss this general and important problem later in more detail, but next we will address the problem of detecting either two high or two low outliers especially before proceeding to tests for many outliers. Also we will return to Example 3-5 for further consideration relative to the so-far-retained value of 1.01.

*In Example 3-5 our resulting or overall significance level turns out to be very close to 90% and is not 95%.

3-5.4 SIGNIFICANCE TESTS FOR THE TWO HIGHEST OR THE TWO LOWEST OBSERVATIONS

To detect whether the two largest or the two smallest observations are probable outliers, we employ a test provided by Grubbs (Refs. 9, 10, 11, and 12). This test is based on the ratio of the sample SS when the two doubtful values (two highest or two lowest) are *omitted* to the total sample SS when the two doubtful values are *included*. If simplicity in calculation is the prime requirement, the Dixon type of test (par. 3-5.2)—actually omitting one observation in the sample—might be used for this case also. In illustrating the test procedure, we will apply the theory to two examples.

Example 3-6:

In a comparison of strength of various plastic materials, one characteristic studied was the percentage of elongation at break. Before comparison of the average elongation of the several materials, it seems desirable to isolate for further study any pieces of a given material that gave very small elongation at breakage compared with the rest of the pieces in the sample. In such an investigation one might have primary interest only in outliers to the left of the mean for study since very high readings indicate exceeding plasticity—a desirable characteristic.

Ten measurements of percentage of elongation at break made on Material No. 23 are 3.73, 3.59, 3.94, 4.13, 3.04, 2.22, 3.23, 4.05, 4.11, and 2.02.

Arranged in ascending order of magnitude, these measurements are 2.02, 2.22, 3.04, 3.23, 3.59, 3.73, 3.94, 4.05, 4.11, 4.13. The questionable readings are the two lowest, 2.02 and 2.22. We can test these two low readings simultaneously by using the following criterion (Refs. 9, 10, 11, and 12):

$$\frac{S_{1,2}^2}{S^2} = \sum_{i=3}^{n}(x_i - \bar{x}_{1,2})^2 \Big/ \sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{3-38}$$

where for the numerator sum of squares the two lowest observations are omitted and

$$\bar{x}_{1,2} = \sum_{i=3}^{n} x_i/(n-2). \tag{3-39}$$

If we were to test the significance of the two highest observations, clearly, the largest and next to largest observations only would be truncated. See the equations at the bottom of Table 3-4.

For the 10 measurements the denominator S^2 of Eq. 3-38 is

$$S^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}{n} \tag{3-40}$$

$$= \frac{10(121.3594) - (34.06)^2}{10} = 5.351$$

and for the truncated sample, using eight measurements,

$$S_{1,2}^2 = \sum_{i=3}^{n}(x_i - \bar{x}_{1,2})^2 = \frac{(n-2)\sum_{i=3}^{n} x_i^2 - (\sum_{i=3}^{n} x_i)^2}{n-2} \tag{3-41}$$

$$= \frac{8(112.3506) - (29.82)^2}{8} = 1.197.$$

Thus we find by Eq. 3-38

$$\frac{S_{1,2}^2}{S^2} = \frac{1.197}{5.351} = 0.224.$$

TABLE 3-4

CRITICAL VALUES FOR $S^2_{n-1,n}/S^2$ OR $S^2_{1,2}/S^2$ FOR SIMULTANEOUSLY TESTING THE TWO LARGEST OR TWO SMALLEST OBSERVATIONS (Ref. 13)*

No. of Obs. n	Lower 0.1% Sig. Level	Lower 0.5% Sig. Level	Lower 1% Sig. Level	Lower 2.5% Sig. Level	Lower 5% Sig. Level	Lower 10% Sig. Level
4	0.0000	0.0000	0.0000	0.0002	0.0008	0.0031
5	0.0003	0.0018	0.0035	0.0090	0.0183	0.0376
6	0.0039	0.0116	0.0186	0.0349	0.0564	0.0920
7	0.0135	0.0308	0.0440	0.0708	0.1020	0.1479
8	0.0290	0.0563	0.0750	0.1101	0.1478	0.1994
9	0.0489	0.0851	0.1082	0.1492	0.1909	0.2454
10	0.0714	0.1150	0.1414	0.1864	0.2305	0.2863
11	0.0953	0.1448	0.1736	0.2213	0.2667	0.3227
12	0.1198	0.1738	0.2043	0.2537	0.2996	0.3552
13	0.1441	0.2016	0.2333	0.2836	0.3295	0.3843
14	0.1680	0.2280	0.2605	0.3112	0.3568	0.4106
15	0.1912	0.2530	0.2859	0.3367	0.3818	0.4345
16	0.2136	0.2767	0.3098	0.3603	0.4048	0.4562
17	0.2350	0.2990	0.3321	0.3822	0.4259	0.4761
18	0.2556	0.3200	0.3530	0.4025	0.4455	0.4944
19	0.2752	0.3398	0.3725	0.4214	0.4636	0.5113
20	0.2939	0.3585	0.3909	0.4391	0.4804	0.5270
21	0.3118	0.3761	0.4082	0.4556	0.4961	0.5415
22	0.3288	0.3927	0.4245	0.4711	0.5107	0.5550
23	0.3450	0.4085	0.4398	0.4857	0.5244	0.5677
24	0.3605	0.4234	0.4543	0.4994	0.5373	0.5795
25	0.3752	0.4376	0.4680	0.5123	0.5495	0.5906
26	0.3893	0.4510	0.4810	0.5245	0.5609	0.6011
27	0.4027	0.4638	0.4933	0.5360	0.5717	0.6110
28	0.4156	0.4759	0.5050	0.5470	0.5819	0.6203
29	0.4279	0.4875	0.5162	0.5574	0.5916	0.6292
30	0.4397	0.4985	0.5268	0.5672	0.6008	0.6375
31	0.4510	0.5091	0.5369	0.5766	0.6095	0.6455
32	0.4618	0.5192	0.5465	0.5856	0.6178	0.6530
33	0.4722	0.5288	0.5557	0.5941	0.6257	0.6602
34	0.4821	0.5381	0.5646	0.6023	0.6333	0.6671
35	0.4917	0.5469	0.5730	0.6101	0.6405	0.6737
36	0.5009	0.5554	0.5811	0.6175	0.6474	0.6800

(cont'd on next page)

$$S^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 \; ; \; \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \; ; \; x_1 \le x_2 \le \cdots \le x_n$$

$$S^2_{1,2} = \sum_{i=2}^{n}(x_i - \bar{x}_{1,2})^2 \; ; \; \bar{x}_{1,2} = \frac{1}{n-2}\sum_{i=2}^{n} x_i$$

$$S^2_{n-1,n} = \sum_{i=1}^{n-2}(x_i - \bar{x}_{n-1,n})^2 \; ; \; \bar{x}_{n-1,n} = \frac{1}{n-2}\sum_{i=1}^{n-2} x_i$$

*A calculated ratio less than the appropriate critical ratio in this table calls for rejection of the null hypothesis.

TABLE 3-4 (cont'd)

No. of Obs. *n*	Lower 0.1% Sig. Level	Lower 0.5% Sig. Level	Lower 1% Sig. Level	Lower 2.5% Sig. Level	Lower 5% Sig. Level	Lower 10% Sig. Level
37	0.5098	0.5636	0.5889	0.6247	0.6541	0.6860
38	0.5184	0.5714	0.5963	0.6316	0.6604	0.6917
39	0.5266	0.5789	0.6035	0.6382	0.6665	0.6972
40	0.5345	0.5862	0.6104	0.6445	0.6724	0.7025
41	0.5422	0.5932	0.6170	0.6506	0.6780	0.7076
42	0.5496	0.5999	0.6234	0.6565	0.6834	0.7125
43	0.5568	0.6064	0.6296	0.6621	0.6886	0.7172
44	0.5637	0.6127	0.6355	0.6676	0.6936	0.7218
45	0.5704	0.6188	0.6412	0.6728	0.6985	0.7261
46	0.5768	0.6246	0.6468	0.6779	0.7032	0.7304
47	0.5831	0.6303	0.6521	0.6828	0.7077	0.7345
48	0.5892	0.6358	0.6573	0.6876	0.7120	0.7384
49	0.5951	0.6411	0.6623	0.6921	0.7163	0.7422
50	0.6008	0.6462	0.6672	0.6966	0.7203	0.7459
51	0.6063	0.6512	0.6719	0.7009	0.7243	0.7495
52	0.6117	0.6560	0.6765	0.7051	0.7281	0.7529
53	0.6169	0.6607	0.6809	0.7091	0.7319	0.7563
54	0.6220	0.6653	0.6852	0.7130	0.7355	0.7595
55	0.6269	0.6697	0.6894	0.7168	0.7390	0.7627
56	0.6317	0.6740	0.6934	0.7205	0.7424	0.7658
57	0.6364	0.6782	0.6974	0.7241	0.7456	0.7687
58	0.6410	0.6823	0.7012	0.7276	0.7489	0.7716
59	0.6454	0.6862	0.7049	0.7310	0.7520	0.7744
60	0.6497	0.6901	0.7086	0.7343	0.7550	0.7772
61	0.6539	0.6938	0.7121	0.7375	0.7580	0.7798
62	0.6580	0.6975	0.7155	0.7406	0.7608	0.7824
63	0.6620	0.7010	0.7189	0.7437	0.7636	0.7850
64	0.6658	0.7045	0.7221	0.7467	0.7664	0.7874
65	0.6696	0.7079	0.7253	0.7496	0.7690	0.7898
66	0.6733	0.7112	0.7284	0.7524	0.7716	0.7921
67	0.6770	0.7144	0.7314	0.7551	0.7741	0.7944
68	0.6805	0.7175	0.7344	0.7578	0.7766	0.7966
69	0.6839	0.7206	0.7373	0.7604	0.7790	0.7988
70	0.6873	0.7236	0.7401	0.7630	0.7813	0.8009
71	0.6906	0.7265	0.7429	0.7655	0.7836	0.8030
72	0.6938	0.7294	0.7455	0.7679	0.7859	0.8050
73	0.6970	0.7322	0.7482	0.7703	0.7881	0.8070
74	0.7000	0.7349	0.7507	0.7727	0.7902	0.8089
75	0.7031	0.7376	0.7532	0.7749	0.7923	0.8108
76	0.7060	0.7402	0.7557	0.7772	0.7944	0.8127
77	0.7089	0.7427	0.7581	0.7794	0.7964	0.8145
78	0.7117	0.7453	0.7605	0.7815	0.7983	0.8162
79	0.7145	0.7477	0.7628	0.7836	0.8002	0.8180
80	0.7172	0.7501	0.7650	0.7856	0.8021	0.8197
81	0.7199	0.7525	0.7672	0.7876	0.8040	0.8213
82	0.7225	0.7548	0.7694	0.7896	0.8058	0.8230
83	0.7250	0.7570	0.7715	0.7915	0.8075	0.8245
84	0.7275	0.7592	0.7736	0.7934	0.8093	0.8261
85	0.7300	0.7614	0.7756	0.7953	0.8109	0.8276

(cont'd on next page)

TABLE 3-4 (cont'd)

No. of Obs. *n*	Lower 0.1% Sig. Level	Lower 0.5% Sig. Level	Lower 1% Sig. Level	Lower 2.5% Sig. Level	Lower 5% Sig. Level	Lower 10% Sig. Level
86	0.7324	0.7635	0.7776	0.7971	0.8126	0.8291
87	0.7348	0.7656	0.7796	0.7989	0.8142	0.8306
88	0.7371	0.7677	0.7815	0.8006	0.8158	0.8321
89	0.7394	0.7697	0.7834	0.8023	0.8174	0.8335
90	0.7416	0.7717	0.7853	0.8040	0.8190	0.8349
91	0.7438	0.7736	0.7871	0.8057	0.8205	0.8362
92	0.7459	0.7755	0.7889	0.8073	0.8220	0.8376
93	0.7481	0.7774	0.7906	0.8089	0.8234	0.8389
94	0.7501	0.7792	0.7923	0.8104	0.8248	0.8402
95	0.7522	0.7810	0.7940	0.8120	0.8263	0.8414
96	0.7542	0.7828	0.7957	0.8135	0.8276	0.8427
97	0.7562	0.7845	0.7973	0.8149	0.8290	0.8439
98	0.7581	0.7862	0.7989	0.8164	0.8303	0.8451
99	0.7600	0.7879	0.8005	0.8178	0.8316	0.8463
100	0.7619	0.7896	0.8020	0.8192	0.8329	0.8475
101	0.7637	0.7912	0.8036	0.8206	0.8342	0.8486
102	0.7655	0.7928	0.8051	0.8220	0.8354	0.8497
103	0.7673	0.7944	0.8065	0.8233	0.8367	0.8508
104	0.7691	0.7959	0.8080	0.8246	0.8379	0.8519
105	0.7708	0.7974	0.8094	0.8259	0.8391	0.8530
106	0.7725	0.7989	0.8108	0.8272	0.8402	0.8541
107	0.7742	0.8004	0.8122	0.8284	0.8414	0.8551
108	0.7758	0.8018	0.8136	0.8297	0.8425	0.8563
109	0.7774	0.8033	0.8149	0.8309	0.8436	0.8571
110	0.7790	0.8047	0.8162	0.8321	0.8447	0.8581
111	0.7806	0.8061	0.8175	0.8333	0.8458	0.8591
112	0.7821	0.8074	0.8188	0.8344	0.8469	0.8600
113	0.7837	0.8088	0.8200	0.8356	0.8479	0.8610
114	0.7852	0.8101	0.8213	0.8367	0.8489	0.8619
115	0.7866	0.8114	0.8225	0.8378	0.8500	0.8628
116	0.7881	0.8127	0.8237	0.8389	0.8510	0.8637
117	0.7895	0.8139	0.8249	0.8400	0.8519	0.8646
118	0.7909	0.8152	0.8261	0.8410	0.8529	0.8655
119	0.7923	0.8164	0.8272	0.8421	0.8539	0.8664
120	0.7937	0.8176	0.8284	0.8431	0.8548	0.8672
121	0.7951	0.8188	0.8295	0.8441	0.8557	0.8681
122	0.7964	0.8200	0.8306	0.8451	0.8567	0.8689
123	0.7977	0.8211	0.8317	0.8461	0.8576	0.8697
124	0.7990	0.8223	0.8327	0.8471	0.8585	0.8705
125	0.8003	0.8234	0.8338	0.8480	0.8593	0.8713
126	0.8016	0.8245	0.8348	0.8490	0.8602	0.8721
127	0.8028	0.8256	0.8359	0.8499	0.8611	0.8729
128	0.8041	0.8267	0.8369	0.8508	0.8619	0.8737
129	0.8053	0.8278	0.8379	0.8517	0.8627	0.8744
130	0.8065	0.8288	0.8389	0.8526	0.8636	0.8752
131	0.8077	0.8299	0.8398	0.8535	0.8644	0.8759
132	0.8088	0.8309	0.8408	0.8544	0.8652	0.8766
133	0.8100	0.8319	0.8418	0.8553	0.8660	0.8773
134	0.8111	0.8329	0.8427	0.8561	0.8668	0.8780

(cont'd on next page)

TABLE 3-4 (cont'd)

No. of Obs. n	Lower 0.1% Sig. Level	Lower 0.5% Sig. Level	Lower 1% Sig. Level	Lower 2.5% Sig. Level	Lower 5% Sig. Level	Lower 10% Sig. Level
135	0.8122	0.8339	0.8436	0.8570	0.8675	0.8787
136	0.8134	0.8349	0.8445	0.8578	0.8683	0.8794
137	0.8145	0.8358	0.8454	0.8586	0.8690	0.8801
138	0.8155	0.8368	0.8463	0.8594	0.8698	0.8808
139	0.8166	0.8377	0.8472	0.8602	0.8705	0.8814
140	0.8176	0.8387	0.8481	0.8610	0.8712	0.8821
141	0.8187	0.8396	0.8489	0.8618	0.8720	0.8827
142	0.8197	0.8405	0.8498	0.8625	0.8727	0.8834
143	0.8207	0.8414	0.8506	0.8633	0.8734	0.8840
144	0.8218	0.8423	0.8515	0.8641	0.8741	0.8846
145	0.8227	0.8431	0.8523	0.8648	0.8747	0.8853
146	0.8237	0.8440	0.8531	0.8655	0.8754	0.8859
147	0.8247	0.8449	0.8539	0.8663	0.8761	0.8865
148	0.8256	0.8457	0.8547	0.8670	0.8767	0.8871
149	0.8266	0.8465	0.8555	0.8677	0.8774	0.8877

Reprinted with permission. Copyright © by the American Statistical Association.

From Table 3-4 for $n = 10$, the 5% significance level for $S^2_{1,2}/S^2$ is 0.2305. A calculated ratio less than the appropriate critical ratio in this table calls for rejection of the null hypothesis. Since the calculated value is less than the critical value, we conclude that both 2.02 and 2.22 are outliers.

In a situation such as the one described in this example, where the outliers are to be isolated for further analysis, a significance level as high as 5% or perhaps even 10% would probably be used to get a reasonable number of sample items for additional study. The problem may really be one of economics, and we should therefore use appropriate probability theory as a sensible basis for action.

Kudo (Ref. 19) indicates that if the two outliers are due to a shift in location or level, as compared to the scale s, then the optimum sample criterion for testing should be of the type

$$\min(2\bar{x} - x_i - x_j)/s = (2\bar{x} - x_1 - x_2)/s \tag{3-42}$$

in our Example 3-6.

In Example 3-7 we give an example in ballistics for which short-range rounds may be due to excessive projectile yaw, i.e., some explainable physical meaning.

Example 3-7:

The following ranges (horizontal distances measured in yards from gun muzzle to point of ground impact of a projectile) were obtained in firings from a weapon at a constant angle of elevation and with the same weight of charge of propellant:

4782	4420
4838	4803
4765	4730
4549	4833.

We desire to make a judgment on whether the projectiles exhibit uniformity in ballistic behavior or whether some of the ranges are inconsistent. The doubtful values are the two smallest ranges, 4420 and 4549 yd. For testing these two suspected outliers, the statistic $S^2_{1,2}/S^2$ of Eq. 3-38 and Table 3-4 is probably the best to use.

The distances, arranged in increasing order of yards or magnitude, are

4420	4782
4549	4803
4730	4833
4765	4838.

The value of S^2 from Eq. 3-40 is 158,592. Omission of the two shortest ranges, 4420 and 4549, and recalculation for the remaining SS gives $S^2_{1,2}$ from Eq. 3-41 equal to 8590.8. Thus

$$\frac{S^2_{1,2}}{S^2} = \frac{8590.8}{158,592} = 0.054$$

which is significant at the 0.01 level. (See Table 3-4.) Therefore, it appears highly unlikely that the two shortest ranges—actually occurring from excessive yaw—could have come from the same population as that represented by the other six ranges for the projectiles. It should be noted that the critical values in Table 3-4 for the 1% level of significance are smaller than those for the 5% level. So for this particular test, we should keep in mind that the calculated value is significant if it is less than the chosen critical value.

If simplicity in calculation is desired or if a large number of samples must be examined individually for outliers, the questionable observations may be tested with the application of Dixon's criteria. *Disregarding only the lowest range, 4420*, and reducing the sample size to seven, we test whether the next lowest range, 4549, is outlying. With $n = 7$ we see from Table 3-2 that r_{10} is the appropriate statistic. Renumbering the ranges as x_1 to x_7, beginning with 4549, we find:

$$r_{10} = \frac{x_2 - x_1}{x_7 - x_1} = \frac{4730 - 4549}{4838 - 4549} = 0.626,$$

which is only a little less than the 1% critical value, 0.637, for $n = 7$. So, if the test is being conducted at any significance level greater than a 1% level, we would conclude that 4549 is an outlier. Since the lowest of the original set of ranges, 4420, is even more outlying than the one we have just tested, it can be classified as an outlier without further testing. We note, however, that this test did not use all of the sample observations.

3-5.5 SIGNIFICANCE TEST FOR DETECTING SEVERAL OR MANY OUTLIERS

3-5.5.1 Preliminary Comments

Although the procedures previously given for detecting a single outlier in a sample have been rather widely studied over the years and have been found to possess about as much power as possible, the problem of detecting several outliers appears to call for much more research. In fact, we commented earlier (par. 3-5.3) that in using the ratio of sample range to standard deviation test to judge whether the largest and smallest observations simultaneously are outliers, one invariably finds that a very satisfactory and clear-cut procedure for rejecting the two extreme values or either one of them is not available without further testing. Thus it appears that tests involving possible outliers on both sides of the sample mean may need much additional study; this applies to several outliers on only one side of the sample mean as well. Indeed, this trend of investigation has been followed in recent years by Tietjen and Moore (Ref. 18), Rosner (Ref. 20), Hawkins (Ref. 21), and others. In view of the analytical complexity involved in the overall problem, much of the statistical research in this area must of necessity resort to Monte Carlo-type simulations to obtain answers, at least for the present time.

3-5.5.2 The Tietjen and Moore Tests

For suspected observations on both the high and low sides in the sample and to deal with the situation in which some of $k \geq 2$ suspected "outliers" are larger and some smaller than the remaining values in the sample, Tietjen and Moore (Ref. 18) suggested the type of statistic that follows. Let the ordered sample values be $x_1, x_2, x_3, \ldots, x_n$, and compute the sample mean $\bar{x}$. Then calculate the n absolute residuals r_i

$$r_1 = |x_1 - \bar{x}|,\ r_2 = |x_2 - \bar{x}|,\ \ldots,\ r_n = |x_n - \bar{x}| \tag{3-43}$$

where the sample mean $\bar{x}$ for the whole, original sample is used. Now relabel the original observations x_1, x_2, . . ., x_n as z's in such a manner that z_i is that original observation x whose r_i is the ith ordered (increasing) absolute residual given by Eq. 3-43. This now means that z_1 is that observation x closest to the mean and that z_n is the observation x farthest from the mean. The Tietjen-Moore (Ref. 18) statistic E_k for testing the significance of the k largest residuals is then

$$E_k = \frac{\sum_{i=1}^{n-k} (z_i - \bar{z}_k)^2}{\sum_{i=1}^{n} (z_i - \bar{z})^2} \tag{3-44}$$

where

$$\begin{aligned} \bar{z}_k &= \sum_{i=1}^{n-k} z_i/(n-k) \\ &= \text{mean of the } (n-k) \text{ least extreme observations} \\ \bar{z} &= \text{mean of the full sample.} \end{aligned} \tag{3-45}$$

The null distribution percentage points of E_k for the two-sided Tietjen-Moore significance test (Ref. 18)—computed by Monte Carlo methods on a high-speed electronic calculator—are given in Table 3-5.

Example 3-8:

Apply the Tietjen-Moore test to the data of Example 3-5 to see whether −1.40 and 1.01 are outliers. We find that the total sum of squares of deviations for the entire sample is 4.24964. Omitting −1.40 and 1.01, the suspected two or largest residual "outliers", we find that the sum of squares of deviations for the reduced sample of 13 observations is 1.24089. From Eq. 3-44 the Tietjen-Moore E_2 = 1.24089/4.24964 = 0.292. Using Table 3-5,* we find that this observed E_2 is somewhat smaller than the 5% critical value of 0.317, so that the E_2 test would reject both of the observations, −1.40 and 1.01. Thus we would probably lean toward taking this latter recommendation since the level of significance for the E_2 test is precisely 0.05, whereas that for the double application of tests for a single outlier, as we carried out in Example 3-5, is greater than 0.05 but less than $1 - (0.95)^2 = 0.0975$. Also we will check this decision to reject −1.40 and 1.01 with the aid of the Rosner (Ref. 20) and Hawkins (Ref. 21) tests in Example 3-9 of par. 3-5.5.3.

Tietjen and Moore (Ref. 18) have also developed tests for suspected outliers on only one side of the sample mean. These are referred to as the L_k Tests of Significance, for the k largest sample values suspected, where

$$L_k = \sum_{i=1}^{n-k} (x_i - \bar{x}_k)^2 \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{3-46}$$

and

$$\bar{x}_k = \sum_{i=1}^{n-k} x_i/(n-k). \tag{3-47}$$

A similar, obvious test for the k smallest suspected sample values is also used by Tietjen and Moore by deletion of these k lowest values in the numerator. Note that the Tietjen-Moore L_2 for either the two highest or two lowest sample values is precisely the $S^2_{n,n-1}/S^2$ or $S^2_{1,2}/S^2$ of Grubbs (Refs. 9, 10, 11, and 12), which is discussed in par. 3-5.4. The L_k percentage points of Tietjen and Moore also were calculated by means of Monte Carlo runs on a high-speed computer and are given in Table 3-6†. Again, the columns headed with an ** indicate the agreement of the Tietjen-Moore Monte Carlo simulations with the exact theoretical percentage points calculated by Grubbs in 1950 for L_1 and L_2 only. Theory for $k \geq 3$ apparent-

* If the calculated ratio is less than the appropriate ratio given in Table 3-5, the values are rejected as outliers.
† If the calculated ratio is less than the appropriate ratio given in Table 3-6, the values are rejected as outliers.

TABLE 3-5

CRITICAL VALUES FOR E_k* (Ref. 18)
$\alpha = 0.01$

$n\backslash k$	1	2	3	4	5	6	7	8	9	10
3	0.000									
4	0.004	0.000								
5	0.029	0.002								
6	0.068	0.012	0.001							
7	0.110	0.028	0.006							
8	0.156	0.050	0.014	0.004						
9	0.197	0.078	0.026	0.009						
10	0.235	0.101	0.037	0.013						
11	0.274	0.134	0.064	0.030	0.012					
12	0.311	0.159	0.083	0.042	0.020	0.008				
13	0.337	0.181	0.103	0.056	0.031	0.014				
14	0.374	0.207	0.123	0.072	0.042	0.022	0.012			
15	0.404	0.238	0.146	0.090	0.054	0.032	0.018			
16	0.422	0.263	0.166	0.107	0.068	0.040	0.024	0.014		
17	0.440	0.290	0.188	0.122	0.079	0.052	0.032	0.018		
18	0.459	0.306	0.206	0.141	0.094	0.062	0.041	0.026	0.014	
19	0.484	0.323	0.219	0.156	0.108	0.074	0.050	0.032	0.020	
20	0.499	0.339	0.236	0.170	0.121	0.086	0.058	0.040	0.026	0.017
25	0.571	0.418	0.320	0.245	0.188	0.146	0.110	0.087	0.066	0.050
30	0.624	0.482	0.386	0.308	0.250	0.204	0.166	0.132	0.108	0.087
35	0.669	0.533	0.435	0.364	0.299	0.252	0.211	0.177	0.149	0.124
40	0.704	0.574	0.480	0.408	0.347	0.298	0.258	0.220	0.190	0.164
45	0.728	0.607	0.518	0.446	0.386	0.336	0.294	0.258	0.228	0.200
50	0.748	0.636	0.550	0.482	0.424	0.376	0.334	0.297	0.264	0.235

(cont'd on next page)

*If the calculated ratio is less than the appropriate ratio given in this table, the values are rejected as outliers.

ly has not been worked out and likely would be very difficult although the Monte Carlo values may certainly be trusted for general use. There is no point in checking the outliers found in Examples 3-6 and 3-7 with the Tietjen-Moore L_2 since that test is equivalent to the one already used.

A point in favor of the Tietjen-Moore type tests is that they clearly cut down or even eliminate the need for and use of several, or multiple, outlier tests.

3-5.5.3 The Rosner and Hawkins Multiple Outlier Detection Procedures

While the Tietjen-Moore procedures for detecting outliers in samples have been valuable in many experimental situations, there have been some improvements since the publication of their paper in 1972 (Ref. 18), especially for the E_k procedure and the rankings called for in Eq. 3-43. In fact, one notes from Eq. 3-43 that all of the rankings of the r_i are based on the original sample mean $\bar{x}$ although it seems more intuitively powerful after finding an outlier to delete that observation from any further consideration and proceed to test the remaining sample values. The point is that an outlier used in the calculation of the sample mean, which is always used in the Tietjen-Moore ranking of Ref. 18, might even mask a second outlier and result in the conclusion that this second outlier is an "inlier" or a perfectly acceptable homogeneous value. This apparently is underlying thoughts of Rosner (Ref. 20) and Hawkins (Ref. 21), and indeed Hawkins (Ref. 21) gives an excellent example to point up this difficulty. Hawkins (Ref. 21) suggests consideration of a sample of $n = 10$ items for which the largest observation $x_n = 100$, the next largest or $x_{n-1} = 10$, and the remaining observations of the sample are from $N(0,1)$, i.e., a normal universe with

TABLE 3-5 (cont'd)
$\alpha = 0.05$*

$n\backslash k$	1	1**	2	3	4	5	6	7	8	9	10
3	0.001	0.001									
4	0.025	0.025	0.001								
5	0.081	0.081	0.010								
6	0.146	0.145	0.034	0.004							
7	0.208	0.207	0.065	0.016							
8	0.265	0.262	0.099	0.034	0.010						
9	0.314	0.310	0.137	0.057	0.021						
10	0.356	0.352	0.172	0.083	0.037	0.014					
11	0.386	0.390	0.204	0.107	0.055	0.026					
12	0.424	0.423	0.234	0.133	0.073	0.039	0.018				
13	0.455	0.453	0.262	0.156	0.092	0.053	0.028				
14	0.484	0.479	0.293	0.179	0.112	0.068	0.039	0.021			
15	0.509	0.503	0.317	0.206	0.134	0.084	0.052	0.030			
16	0.526	0.525	0.340	0.227	0.153	0.102	0.067	0.041	0.024		
17	0.544	0.544	0.362	0.248	0.170	0.116	0.078	0.050	0.032		
18	0.562	0.562	0.382	0.267	0.187	0.132	0.091	0.062	0.041	0.026	
19	0.581	0.579	0.398	0.287	0.203	0.146	0.105	0.074	0.050	0.033	
20	0.597	0.594	0.416	0.302	0.221	0.163	0.119	0.085	0.059	0.041	0.028
25	0.652	0.654	0.493	0.381	0.298	0.236	0.186	0.146	0.114	0.089	0.068
30	0.698		0.549	0.443	0.364	0.298	0.246	0.203	0.166	0.137	0.112
35	0.732		0.596	0.495	0.417	0.351	0.298	0.254	0.214	0.181	0.154
40	0.758		0.629	0.534	0.458	0.395	0.343	0.297	0.259	0.223	0.195
45	0.778		0.658	0.567	0.492	0.433	0.381	0.337	0.299	0.263	0.233
50	0.797		0.684	0.599	0.529	0.468	0.417	0.373	0.334	0.299	0.268

(cont'd on next page)

*If the calculated ratio is less than the appropriate ratio given in this table, the values are rejected as outliers.

**From Grubbs, Table I, Ref. 9. Note in this connection that the Tietjen-Moore Monte Carlo values of Ref. 18 check the Grubbs theoretical 0.05 probability levels of Ref. 9.

mean of zero and standard deviation of unity. Hawkins then points out that the two largest values, 100 and 10 are truly outliers, whereas the original sample mean $\bar{x}$ is about 11, which perhaps brands the value 10 as an inlier. That is to say, the Tietjen-Moore tests (E or L) would test $x_n = x_{10} = 100$ correctly but would sometimes miss the outlier $x_{n-1} = x_9 = 10$ by finally testing the algebraically largest of the remaining eight values, one or more of which on occasion would exceed the $x_9 = 10$.

In 1975 Rosner (Ref. 20) made a rather significant advance in the problem of detecting multiple outliers in a sample by attempting to get away from testing for a prefixed or specified number of outliers, i.e., developing a more flexible procedure to detect from one to k outliers and yet keep the significance level fixed at α. The chief advantage of the Rosner approach is that it should be powerful enough to detect any number of outliers up to $[pn]$, where p is some fraction of the total sample size, and not lose much power against an alternative of a specified number of outliers. Conversely, as Rosner points out, any outlier detection test that is geared to finding a specific number of aberrant values can be much less powerful in detecting any other number of deviant sample observations. Indeed, the number of outliers to expect in advance is hardly ever known, and there is the obvious need to apply a routine rule for any possible number of outliers that may actually be in the sample rather than first trying to guess the correct number by simply observing the data and then using a rule that is good against that particular number of outliers. This means that the Type I error, or α, must be controlled at its present level throughout the sequential testing for as many as k outliers. Rosner's procedure (Refs. 20 and 22) is to employ a set of R statistics, or "RST" multiple outlier tests, as he calls them. Rosner (Refs. 20 and 22) decides in advance that he will test a sample of observations for up to as many as k outliers. The number k is in fact rather arbitrary and

TABLE 3-5 (cont'd)
$\alpha = 0.10$*

$n\backslash k$	1	1**	2	3	4	5	6	7	8	9	10
3	0.003	0.003									
4	0.050	0.049	0.002								
5	0.127	0.127	0.022								
6	0.204	0.203	0.056	0.009							
7	0.268	0.270	0.094	0.027							
8	0.328	0.326	0.137	0.053	0.016						
9	0.377	0.374	0.175	0.080	0.032						
10	0.420	0.415	0.214	0.108	0.052	0.022					
11	0.449	0.451	0.250	0.138	0.073	0.036					
12	0.485	0.482	0.278	0.162	0.094	0.052	0.026				
13	0.510	0.510	0.309	0.189	0.116	0.068	0.038				
14	0.538	0.534	0.337	0.216	0.138	0.086	0.052	0.029			
15	0.558	0.556	0.360	0.240	0.160	0.105	0.067	0.040			
16	0.578	0.576	0.384	0.263	0.182	0.122	0.082	0.053	0.032		
17	0.594	0.593	0.406	0.284	0.198	0.140	0.095	0.064	0.042		
18	0.610	0.610	0.424	0.304	0.217	0.156	0.110	0.076	0.051	0.034	
19	0.629	0.624	0.442	0.322	0.234	0.172	0.124	0.089	0.062	0.042	
20	0.644	0.638	0.460	0.338	0.252	0.188	0.138	0.102	0.072	0.051	0.035
25	0.693	0.692	0.528	0.417	0.331	0.264	0.210	0.168	0.132	0.103	0.080
30	0.730		0.582	0.475	0.391	0.325	0.270	0.224	0.186	0.154	0.126
35	0.763		0.624	0.523	0.443	0.379	0.324	0.276	0.236	0.202	0.172
40	0.784		0.657	0.562	0.486	0.422	0.367	0.320	0.278	0.243	0.212
45	0.803		0.684	0.593	0.522	0.459	0.406	0.360	0.320	0.284	0.252
50	0.820		0.708	0.622	0.552	0.492	0.440	0.396	0.355	0.319	0.287

*If the calculated ratio is less than the appropriate ratio given in this table, the values are rejected as outliers.

**From Grubbs, Table I, Ref. 9. Note in this connection that the Tietjen-Moore Monte Carlo values of Ref. 18 check the Grubbs theoretical 0.10 probability levels of Ref. 9.

Reprinted with permission. Copyright © by the American Statistical Association.

is used to "lop off" or trim the k largest and k smallest observations from the sample so that only an inner sample having no outliers remains and provides a "trimmed" reference sample for a "safe" mean and sigma. He then calculates the trimmed mean a and trimmed variance b^2 for the remaining sample values, or the inliers, which are

$$a = \sum_{i=k+1}^{n-k} x_i/(n-2k) = \text{trimmed mean} \tag{3-48}$$

$$b^2 = \sum_{i=k+1}^{n-k} (x_i - a)^2/(n-2k-1) = \text{trimmed variance.} \tag{3-49}$$

Rosner (Refs. 20 and 22) then calculates the largest studentized residual in absolute value R_1 for the entire sample, but he uses a and b instead of the $\bar{x}$ and s of the whole sample. Thus the observed value of R_1 is calculated as

$$R_1 = \max_{x_i} |x_i - a|/b = |x^{(1)} - a|/b \tag{3-50}$$

where

$x^{(1)}$ = particular value that makes R_1 a maximum.

The calculated value of R_1 is tested statistically against a percentage point or probability level computed by Rosner for R_1 by Monte Carlo methods. Thus the value $x^{(1)}$, which will turn out to be the farthest value from the trimmed mean, is then branded either an outlier or not, but if judged an outlier, it is not considered in the computation of the next studentized residual R_2.

TABLE 3-6

CRITICAL VALUES FOR L_k^* (Ref. 18)
$\alpha = 0.01$

$n\backslash k$	1	1**	2	2***	3	4	5	6	7	8	9	10
3	0.000	0.000										
4	0.011	0.010	0.000	0.000								
5	0.045	0.044	0.004	0.004								
6	0.091	0.093	0.021	0.019	0.002							
7	0.148	0.145	0.047	0.044	0.010							
8	0.202	0.195	0.076	0.075	0.028	0.008						
9	0.235	0.241	0.112	0.108	0.048	0.018						
10	0.280	0.283	0.142	0.141	0.070	0.032	0.012					
11	0.327	0.321	0.178	0.174	0.098	0.052	0.026					
12	0.371	0.355	0.208	0.204	0.120	0.070	0.038	0.019				
13	0.400	0.386	0.233	0.233	0.147	0.094	0.056	0.033				
14	0.424	0.414	0.267	0.261	0.172	0.113	0.072	0.042	0.027			
15	0.450	0.440	0.294	0.286	0.194	0.132	0.090	0.057	0.037			
16	0.473	0.463	0.311	0.310	0.219	0.151	0.108	0.072	0.049	0.030		
17	0.480	0.485	0.338	0.332	0.237	0.171	0.126	0.091	0.064	0.044		
18	0.502	0.504	0.358	0.353	0.260	0.192	0.140	0.104	0.076	0.053	0.036	
19	0.508	0.522	0.366	0.373	0.272	0.201	0.154	0.118	0.088	0.064	0.046	
20	0.533	0.539	0.387	0.391	0.300	0.231	0.175	0.136	0.104	0.078	0.058	0.042
25	0.607		0.468		0.377	0.308	0.246	0.204	0.168	0.144	0.112	0.092
30	0.650		0.527		0.434	0.369	0.312	0.268	0.229	0.196	0.166	0.142
35	0.690		0.573		0.484	0.418	0.364	0.321	0.282	0.250	0.220	0.194
40	0.722		0.610		0.522	0.460	0.408	0.364	0.324	0.292	0.262	0.234
45	0.745		0.641		0.558	0.498	0.444	0.399	0.361	0.328	0.296	0.270
50	0.768		0.667		0.592	0.531	0.483	0.438	0.400	0.368	0.336	0.308

(cont'd on next page)

*If the calculated ratio is less than the appropriate ratio given in this table, the values are rejected as outliers.

**From Grubbs, Table I, Ref. 9. Use instead of Tietjen-Moore Monte Carlo values.

***From Grubbs, Table V, Ref. 9. Use instead of Tietjen-Moore Monte Carlo values.

If $x^{(1)}$ is discardable, the same trimmed mean a and trimmed standard deviation b are used to calculate R_2, the next RST given by

$$R_2 = \max_{x_i} |x_i - a|/b = |x^{(2)} - a|/b \tag{3-51}$$

where

$x^{(2)}$ = particular subsample value that makes R_2 a maximum.

The sample values tested do not include $x^{(1)}$. The process is continued through R_3, etc., to R_k, stopping there or before. In effect, therefore, the Rosner outlier test procedure is sequential in nature and calls for multiple significance tests. This means that a series of calculations is necessary, and the determination of an outlier has to be made at each testing stage.

Rosner (Ref. 20) works with the marginal distributions of R_1, R_2, . . ., and R_k to determine specifically tne values of β, the correct probability level at each stage, and the percent points $\lambda_1(\beta)$, $\lambda_2(\beta)$, . . ., $\lambda_k(\beta)$ such that

$$Pr\,[R_i > \lambda_i(\beta)] = \beta, \quad i = 1, \ldots, k \tag{3-52}$$

TABLE 3-6 (cont'd)
$\alpha = 0.025$*

$n\backslash k$	1	1**	2	2***	3	4	5	6	7	8	9	10
3	0.001	0.001	0.000	0.000								
4	0.025	0.025	0.000	0.000								
5	0.084	0.081	0.011	0.009								
6	0.146	0.145	0.034	0.035	0.005							
7	0.209	0.207	0.076	0.071	0.021							
8	0.262	0.262	0.115	0.110	0.045	0.013						
9	0.308	0.310	0.150	0.149	0.073	0.030						
10	0.350	0.353	0.188	0.187	0.100	0.052	0.023					
11	0.366	0.390	0.225	0.221	0.129	0.074	0.040					
12	0.440	0.423	0.268	0.254	0.162	0.096	0.057	0.031				
13	0.462	0.453	0.292	0.284	0.184	0.122	0.077	0.047				
14	0.493	0.479	0.317	0.311	0.214	0.145	0.098	0.063	0.038			
15	0.498	0.503	0.341	0.337	0.239	0.167	0.111	0.078	0.051			
16	0.537	0.525	0.372	0.360	0.261	0.185	0.137	0.096	0.065	0.045		
17	0.552	0.544	0.388	0.382	0.282	0.208	0.156	0.117	0.082	0.058		
18	0.570	0.562	0.406	0.403	0.299	0.226	0.171	0.129	0.095	0.068	0.048	
19	0.573	0.579	0.416	0.421	0.311	0.243	0.189	0.145	0.108	0.080	0.059	
20	0.595	0.594	0.442	0.439	0.341	0.265	0.209	0.165	0.128	0.098	0.073	0.054
25	0.654		0.512		0.417	0.342	0.282	0.233	0.192	0.159	0.132	0.113
30	0.699		0.567		0.479	0.408	0.352	0.302	0.261	0.226	0.193	0.165
35	0.732		0.610		0.527	0.455	0.398	0.348	0.308	0.274	0.242	0.213
40	0.755		0.644		0.561	0.491	0.433	0.387	0.348	0.314	0.283	0.257
45	0.773		0.667		0.592	0.529	0.473	0.430	0.391	0.356	0.325	0.295
50	0.796		0.697		0.622	0.559	0.510	0.466	0.428	0.392	0.363	0.334

(cont'd on next page)

*If the calculated ratio is less than the appropriate ratio given in this table, the values are rejected as outliers.
**From Grubbs, Table I, Ref. 9. Use instead of Tietjen-Moore Monte Carlo values.
***From Grubbs, Table V, Ref. 9. Use instead of Tietjen-Moore Monte Carlo values.

and the union U of all these sets gives also

$$Pr\left\{\bigcup_{i=1}^{k}[R_i > \lambda_i(\beta)]\right\} = \alpha. \tag{3-53}$$

Rosner (Ref. 22) then establishes the percentage points $\lambda_i(\beta)$ for the R_i with increasing $i = 1, 2, 3, 4$, etc. Such investigations, including especially the power of the detection procedures to reject false null hypotheses, must be made through the means of Monte Carlo-type simulations, which aided Rosner in coming to the following conclusions. He found that the one-outlier detection procedures were slightly more powerful in detecting a single outlier than the several or many outlier detection rules were. However, such advantage seems to be rather slight when compared with the substantial increase in power obtained for the alternative of two or more outliers, particularly when the outliers are on the same side of the mean. The greatest improvement in power for the many outlier detection rules was for the case of multiple outliers on one side of the sample mean, as in the example of Hawkins previously cited. Rosner (Ref. 20) therefore concludes positively that the many outlier detection procedures are preferable to their one-outlier counterparts, particularly if all of the outliers are on the same side of the sample mean. Moreover, by using a multiple outlier detection procedure, instead of a single outlier rule, one tends to give up some power against the alternative of one actual outlier (probably at most 10% depending on the alternative), however, one gains much more power against alternatives of several outliers, and as much as 50% for alternatives where the real outliers are on the same side of the sample mean. Even though one has to give

TABLE 3-6 (cont'd)
$\alpha = 0.05$*

$n\backslash k$	1	1**	2	2***	3	4	5	6	7	8	9	10
3	0.003	0.003										
4	0.051	0.049	0.001	0.001								
5	0.125	0.127	0.018	0.018								
6	0.203	0.203	0.055	0.057	0.010							
7	0.273	0.270	0.106	0.102	0.032							
8	0.326	0.326	0.146	0.148	0.064	0.022						
9	0.372	0.374	0.194	0.191	0.099	0.045						
10	0.418	0.415	0.233	0.230	0.129	0.070	0.034					
11	0.454	0.451	0.270	0.267	0.162	0.098	0.054					
12	0.489	0.482	0.305	0.300	0.196	0.125	0.076	0.042				
13	0.517	0.510	0.337	0.330	0.224	0.150	0.098	0.060				
14	0.540	0.534	0.363	0.357	0.250	0.174	0.122	0.079	0.050			
15	0.556	0.556	0.387	0.382	0.276	0.197	0.140	0.097	0.066			
16	0.575	0.576	0.410	0.405	0.300	0.219	0.159	0.115	0.082	0.055		
17	0.594	0.593	0.427	0.426	0.322	0.240	0.181	0.136	0.100	0.072		
18	0.608	0.610	0.447	0.446	0.337	0.259	0.200	0.154	0.116	0.086	0.062	
19	0.624	0.624	0.462	0.464	0.354	0.277	0.209	0.168	0.130	0.099	0.074	
20	0.639	0.638	0.484	0.480	0.377	0.299	0.238	0.188	0.150	0.115	0.088	0.066
25	0.696	0.692	0.550		0.450	0.374	0.312	0.262	0.222	0.184	0.154	0.126
30	0.730		0.601		0.506	0.434	0.376	0.327	0.283	0.245	0.212	0.183
35	0.762		0.641		0.554	0.482	0.424	0.376	0.334	0.297	0.264	0.235
40	0.784		0.673		0.588	0.523	0.468	0.421	0.378	0.342	0.310	0.280
45	0.802		0.698		0.618	0.556	0.502	0.456	0.417	0.382	0.350	0.320
50	0.820		0.720		0.646	0.588	0.535	0.490	0.450	0.414	0.383	0.356

(cont'd on next page)

*If the calculated ratio is less than the appropriate ratio given in this table, the values are rejected as outliers.
**From Grubbs, Table I, Ref. 9. Use instead of Tietjen-Moore Monte Carlo values.
***From Grubbs, Table V, Ref. 9. Use instead of Tietjen-Moore Monte Carlo values.

up some power against the alternative of two outliers when a multiple outlier procedure is used, the advantage is that one does not have to declare two outliers when in fact only one outlier is actually present; this reduces the number of false positives. Rosner appeared to prefer the extreme studentized deviate (ESD) procedure of Eqs. 3-50 and 3-51 over other rejection rules he studied because they seemed to be the best and were "computationally reasonable". By using Monte Carlo methods, Rosner (Ref. 22) found the $\lambda_i(\beta)$ for certain sample sizes and the maximum number k of outliers suspected in the sample, and we give these in Tables 3-7, 3-8, and 3-9. Example 3-9 illustrates the Rosner procedure.

Example 3-9:

Return to the data of Example 3-5 for the 15 observations concerning the semidiameter measurements of Venus and apply Rosner's outlier test procedure to determine whether -1.40 and 1.01 both should be branded as outliers.

The 15 observations ranked in increasing order are -1.40, -0.44, -0.30, -0.24, -0.22, -0.13, -0.05, 0.06, 0.10, 0.18, 0.20, 0.39, 0.48, 0.63, and 1.01. Now we suspect that at most -1.40 and 1.01 are outliers, so that we may as well put $k = 2$, and censor the two lowest values, -1.40 and -0.44, and the two highest values, 0.63 and 1.01, for the purpose of calculating the trimmed mean a and trimmed standard deviation b. We use -0.30, -0.24, -0.22, -0.13, -0.05, 0.06, 0.10, 0.18, 0.20, 0.39, and 0.48 in Eqs. 3-48 and 3-49 to get

$$a = 0.04273 \text{ and } b = 0.2576.$$

TABLE 3-6 (cont'd)
$\alpha = 0.10$*

$n\backslash k$	1	1**	2	2***	3	4	5	6	7	8	9	10
3	0.011	0.011										
4	0.098	0.098	0.003	0.003								
5	0.200	0.199	0.038	0.038								
6	0.280	0.283	0.091	0.092	0.020							
7	0.348	0.350	0.148	0.148	0.056							
8	0.404	0.405	0.200	0.199	0.095	0.038						
9	0.448	0.450	0.248	0.245	0.134	0.068						
10	0.490	0.488	0.287	0.286	0.170	0.098	0.051					
11	0.526	0.520	0.326	0.323	0.208	0.128	0.074					
12	0.555	0.548	0.361	0.355	0.240	0.159	0.103	0.062				
13	0.578	0.573	0.388	0.384	0.270	0.186	0.126	0.082				
14	0.600	0.594	0.416	0.411	0.298	0.212	0.150	0.104	0.068			
15	0.611	0.613	0.436	0.435	0.322	0.236	0.172	0.124	0.086			
16	0.631	0.631	0.458	0.456	0.342	0.260	0.194	0.144	0.104	0.073		
17	0.648	0.646	0.478	0.476	0.364	0.282	0.216	0.165	0.125	0.092		
18	0.661	0.660	0.496	0.494	0.384	0.302	0.236	0.184	0.142	0.108	0.080	
19	0.676	0.673	0.510	0.511	0.398	0.316	0.251	0.199	0.158	0.124	0.094	
20	0.688	0.685	0.530	0.527	0.420	0.339	0.273	0.220	0.176	0.140	0.110	0.085
25	0.732	0.732	0.591		0.489	0.412	0.350	0.296	0.251	0.213	0.180	0.152
30	0.766		0.637		0.523	0.472	0.411	0.359	0.316	0.276	0.240	0.210
35	0.792		0.674		0.586	0.516	0.458	0.410	0.365	0.328	0.294	0.262
40	0.812		0.702		0.622	0.554	0.499	0.451	0.408	0.372	0.338	0.307
45	0.826		0.726		0.648	0.586	0.533	0.488	0.447	0.410	0.378	0.348
50	0.840		0.746		0.673	0.614	0.562	0.518	0.477	0.442	0.410	0.380

*If the calculated ratio is less than the appropriate ratio given in this table, the values are rejected as outliers.

**From Grubbs, Table I, Ref. 9. Use instead of Tietjen-Moore Monte Carlo values.

***From Grubbs, Table V, Ref. 9. Use instead of Tietjen-Moore Monte Carlo values.

Reprinted with permission. Copyright © for portion of table by American Statistical Association. Copyright © for remainder of table by Institute of Mathematical Statistics.

Hence proceeding to apply Eqs. 3-50 and 3-51, one finds that

$$R_1 = |-1.40 - 0.0427|/0.2576 = 5.60$$

and

$$R_2 = |1.01 - 0.0427|/0.2576 = 3.76.$$

From Rosner's Table 3-7 for $n = 15$, we find that neither -1.40 nor 1.01 are rejectable at the 5% level, but only -1.40 is an outlier at the 10% level! This is somewhat of a surprise because the Tietjen-Moore test rejected both -1.40 and 1.01. Hence we will next examine Hawkins' test and review this matter again in Example 3-10.

In an extended study of the problem of multiple outliers, Hawkins (Ref. 21) points out that Rosner (Ref. 20) apparently noticed the masking-type defect in the widely used Tietjen-Moore E_k statistic (Ref. 18) but did not actually highlight the finding specifically. Hawkins (Ref. 21) also states that the rationale behind the Rosner scheme matches that which one would use intuitively. When trying to decide whether a particular observation is an outlier, one should delete from the sample all observations already concluded to be outliers. Also this is in consonance with the ideas behind the $S^2_{1,2}/S^2$ outlier type tests of Grubbs (Ref. 9). Hawkins also points out that the Rosner ranking procedure leads for any number k of outliers to a set of retained inliers with minimum variance as is the case for likelihood ratio test statistics. Finally,

TABLE 3-7
PERCENTAGE POINTS OF ROSNER'S RST MANY OUTLIER TEST STATISTICS R_1 AND R_2 (Ref. 22)*

$n = 10(5)20(10)50(25)100$ and $k = 2$

n		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
10	R_1	7.35 ± 0.102	8.90 ± 0.146	13.38 ± 0.748
	R_2	4.92 ± 0.067	5.92 ± 0.103	9.13 ± 0.407
15	R_1	5.28 ± 0.63	6.01 ± 0.056	8.10 ± 0.208
	R_2	3.84 ± 0.045	4.31 ± 0.060	5.39 ± 0.134
20	R_1	4.64 ± 0.043	5.18 ± 0.053	6.47 ± 0.182
	R_2	3.50 ± 0.024	3.81 ± 0.032	4.70 ± 0.095
30	R_1	4.26 ± 0.027	4.62 ± 0.037	5.51 ± 0.108
	R_2	3.31 ± 0.021	3.57 ± 0.017	4.15 ± 0.053
40	R_1	4.04 ± 0.019	4.41 ± 0.033	5.26 ± 0.047
	R_2	3.23 ± 0.017	3.43 ± 0.030	3.92 ± 0.042
50	R_1	3.98 ± 0.013	4.25 ± 0.019	4.98 ± 0.081
	R_2	3.20 ± 0.011	3.39 ± 0.022	3.80 ± 0.047
75	R_1	3.89 ± 0.016	4.16 ± 0.016	4.77 ± 0.074
	R_2	3.19 ± 0.013	3.37 ± 0.029	3.72 ± 0.038
100	R_1	3.83 ± 0.016	4.09 ± 0.027	4.66 ± 0.088
	R_2	3.20 ± 0.012	3.34 ± 0.0076	3.74 ± 0.037

The $\pm$ values are standard errors.

This is Table 1 of Rosner.

*For later tables associated with outlier procedures, see also Jain (Ref. 23).

Reprinted with permission. Copyright © by the American Statistical Association.

Hawkins (Ref. 21) allows for an extension of the family of statistics to include the considerations of Paulson (Ref. 24) and Quesenberry and David (Ref. 25) who provided for the case in which there may also be available some additional information on the underlying standard deviation σ in the form of previous or extraneous data to the immediate problem at hand. In such case, an extraneous sum of squares would provide an independent estimator of σ^2 in the form of

$$U^2/\sigma^2 = \chi^2(\nu) \quad (3\text{-}54)$$
$$= \text{chi-square with } \nu \text{ df}$$

where

U^2 = an independent sum of squares to estimate the variance

σ^2 = estimated population variance.

Hawkins then defines the extended statistic E_k^* as

$$E_k^* = (S_k^{2*} + U)/(S + U) \quad (3\text{-}55)$$

TABLE 3-8
PERCENTAGE POINTS OF ROSNER'S RST MANY OUTLIER TEST STATISTICS R_1, R_2, AND R_3 (Ref. 22)

$n = 20(10)50(25)100$ and $k = 3$

n		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
20	R_1	5.91 ± 0.059	6.60 ± 0.079	8.19 ± 0.137
	R_2	4.50 ± 0.047	5.06 ± 0.052	6.34 ± 0.151
	R_3	3.73 ± 0.037	4.16 ± 0.046	5.22 ± 0.098
30	R_1	5.07 ± 0.037	5.60 ± 0.063	6.88 ± 0.093
	R_2	3.93 ± 0.028	4.32 ± 0.037	5.09 ± 0.121
	R_3	3.35 ± 0.016	3.62 ± 0.039	4 .27± 0.076
40	R_1	4.60 ± 0.037	5.06 ± 0.040	6.05 ± 0.103
	R_2	3.68 ± 0.021	3.92 ± 0.021	4.53 ± 0.051
	R_3	3.20 ± 0.016	3.41 ± 0.024	3.82 ± 0.063
50	R_1	4.43 ± 0.033	4.76 ± 0.049	5.68 ± 0.038
	R_2	3.60 ± 0.014	3.82 ± 0.018	4.55 ± 0.086
	R_3	3.14 ± 0.019	3.30 ± 0.014	3.77 ± 0.047
75	R_1	4.18 ± 0.024	4.46 ± 0.034	5.10 ± 0.036
	R_2	3.47 ± 0.013	3.67 ± 0.019	4.10 ± 0.040
	R_3	3.08 ± 0.0096	3.19 ± 0.012	3.57 ± 0.045
100	R_1	4.12 ± 0.019	4.37 ± 0.034	4.98 ± 0.120
	R_2	3.44 ± 0.012	3.60 ± 0.022	3.88 ± 0.039
	R_3	3.10 ± 0.012	3.21 ± 0.016	3.45 ± 0.031

The ± values are standard errors.
This is Table 2 of Rosner.

Reprinted with permission. Copyright © by the American Statistical Association.

where

S_k^{2*} = inlier SS

S^2 = SS for the entire sample

as a suggested test statistic for the presence of k outliers for the additional or past information U on the unknown σ^2. In the event that no external information on σ is available, one simply sets $U = \nu = 0$, and the statistic E_k^* becomes the inlier SS divided by the SS for the entire sample, i.e., the Grubbs (Ref. 9) type test. By a Monte Carlo process Hawkins (Ref. 21) calculates tables of percentage points of the statistic E_k^*; this information is in Table 3-10. It is believed that these new tables of percentage points of Hawkins should be of rather wide application, and Example 3-10 is an example of their use.

Example 3-10:

Consider again the 15 observations on the semidiameter measurements of Venus in Example 3-5 and also Example 3-8, where we used the Tietjen-Moore E_2 test and rejected both the −1.40 and 1.01 observations.

We have, as before, that the inlier sum of squares is 1.2409, and the total sample SS is 4.2496. Hence for $\nu = 0$ there is no difference between the Tietjen-Moore test and that of Hawkins. We note that the 5%

TABLE 3-9
PERCENTAGE POINTS OF ROSNER'S RST MANY OUTLIER TEST STATISTICS
R_1, R_2, R_3 and R_4 (Ref. 22)

$n = 20(10)50(25)100$ and $k = 4$

n		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
20	R_1	7.56 ± 0.083	8.52 ± 0.112	11.70 ± 0.340
	R_2	5.88 ± 0.042	6.53 ± 0.050	8.83 ± 0.263
	R_3	4.91 ± 0.038	5.46 ± 0.064	7.23 ± 0.199
	R_4	4.17 ± 0.035	4.65 ± 0.056	6.03 ± 0.116
30	R_1	5.90 ± 0.030	6.40 ± 0.055	7.65 ± 0.096
	R_2	4.63 ± 0.030	5.01 ± 0.034	5.90 ± 0.094
	R_3	3.95 ± 0.037	4.27 ± 0.049	5.09 ± 0.089
	R_4	3.50 ± 0.024	3.76 ± 0.034	4.53 ± 0.101
40	R_1	5.23 ± 0.036	5.67 ± 0.066	6.85 ± 0.264
	R_2	4.13 ± 0.025	4.47 ± 0.037	5.24 ± 0.087
	R_3	3.60 ± 0.031	3.82 ± 0.030	4.52 ± 0.079
	R_4	3.25 ± 0.020	3.43 ± 0.027	3.99 ± 0.043
50	R_1	4.85 ± 0.036	5.19 ± 0.063	6.18 ± 0.111
	R_2	3.95 ± 0.022	4.18 ± 0.028	4.86 ± 0.082
	R_3	3.46 ± 0.014	3.67 ± 0.019	4.20 ± 0.066
	R_4	3.14 ± 0.0098	3.30 ± 0.021	3.75 ± 0.041
75	R_1	4.55 ± 0.039	4.87 ± 0.060	5.66 ± 0.105
	R_2	3.73 ± 0.022	3.94 ± 0.018	4.41 ± 0.054
	R_3	3.31 ± 0.010	3.47 ± 0.020	3.81 ± 0.021
	R_4	3.04 ± 0.014	3.16 ± 0.019	3.50 ± 0.034
100	R_1	4.43 ± 0.037	4.67 ± 0.034	5.38 ± 0.091
	R_2	3.64 ± 0.016	3.80 ± 0.018	4.28 ± 0.056
	R_3	3.27 ± 0.012	3.39 ± 0.011	3.72 ± 0.037
	R_4	3.03 ± 0.011	3.14 ± 0.012	3.41 ± 0.028

The $\pm$ values are standard errors.
This is Table 3 of Rosner.

Reprinted with permission. Copyright © by the American Statistical Association.

level of E_2^* for $n = 15$ in Table 3-10† is 0.3104, whereas that of Tietjen and Moore in Table 3-5 is 0.317. Note that Hawkins indicates his Monte Carlo calculations are good to perhaps four decimal places. We decide to reject both -1.40 and 1.01 because we believe the sum of squares type test may be superior to the Rosner outlier test. This is our final conclusion for these data.

3-5.5.4 The Skewness and Kurtosis Tests for Outliers

In our account of testing samples for multiple outliers, we should also record some discussion concerning the related work of Ferguson (Refs. 15 and 16). In fact, the use of the skewness and kurtosis coefficients have long been studied as tests of normality and also as a way of screening samples for outliers. We have already mentioned the matter of possible spurious values in the sample being masked by the presence of other anomalous observations since this will have an effect on any significance tests to detect outlying observations. Outlying observations occur due to a shift in level (or mean) or a change in scale

† If the calculated ratio is less than the appropriate ratio given in Table 3-10, the values are rejected as outliers.

TABLE 3-10

PERCENTAGE POINTS OF HAWKINS' E_K* (Ref. 21)†

		df = 0			df = 10			df = 20		
k	$n\backslash\alpha$	0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001
2	5	0.0061	0.0011	0.0001	0.4556	0.3380	0.2205	0.6613	0.5610	0.4433
2	10	0.1640	0.0995	0.0487	0.4920	0.3985	0.2948	0.6461	0.5650	0.4664
2	15	0.3104	0.2319	0.1529	0.5404	0.4592	0.3638	0.6596	0.5890	0.5010
2	20	0.4136	0.3367	0.2509	0.5829	0.5105	0.4222	0.6780	0.6148	0.5345
2	25	0.4886	0.4168	0.3320	0.6186	0.5531	0.4712	0.6964	0.6388	0.5647
2	30	0.5455	0.4792	0.3982	0.6486	0.5887	0.5126	0.7135	0.6606	0.5917
2	40	0.6262	0.5698	0.4977	0.6959	0.6449	0.5782	0.7432	0.6975	0.6369
2	50	0.6810	0.6322	0.5684	0.7314	0.6869	0.6278	0.7676	0.7272	0.6731
2	75	0.7639	0.7277	0.6788	0.7907	0.7569	0.7111	0.8117	0.7804	0.7377
2	100	0.8109	0.7821	0.7428	0.8275	0.8002	0.7628	0.8412	0.8156	0.7803
3	10	0.0743	0.0395	0.0160	0.4000	0.3117	0.2190	0.5800	0.4959	0.3973
3	15	0.1967	0.1389	0.0847	0.4454	0.3684	0.2814	0.5860	0.5141	0.4270
3	20	0.2972	0.2338	0.1661	0.4892	0.4197	0.3376	0.6028	0.5386	0.4591
3	25	0.3758	0.3128	0.2410	0.5278	0.4641	0.3867	0.6217	0.5632	0.4896
3	30	0.4379	0.3776	0.3057	0.5612	0.5024	0.4292	0.6402	0.5863	0.5175
3	40	0.5297	0.4758	0.4084	0.6153	0.5642	0.4988	0.6737	0.6268	0.5657
3	50	0.5942	0.5463	0.4846	0.6570	0.6119	0.5530	0.7021	0.6604	0.6053
3	75	0.6948	0.6579	0.6087	0.7286	0.6936	0.6467	0.7551	0.7223	0.6782
3	100	0.7534	0.7236	0.6832	0.7741	0.7456	0.7067	0.7915	0.7645	0.7275
4	10	0.0299	0.0132	0.0041	0.3360	0.2515	0.1673	0.5357	0.4478	0.3485
4	15	0.1239	0.0819	0.0455	0.3753	0.3022	0.2225	0.5313	0.4583	0.3721
4	20	0.2154	0.1631	0.1099	0.4182	0.3516	0.2752	0.5449	0.4803	0.4019
4	25	0.2923	0.2370	0.1760	0.4575	0.3959	0.3227	0.5631	0.5043	0.4316
4	30	0.3558	0.3006	0.2368	0.4924	0.4350	0.3650	0.5819	0.5277	0.4595
4	40	0.4530	0.4015	0.3383	0.5504	0.4998	0.4360	0.6174	0.5700	0.5090
4	50	0.5235	0.4764	0.4169	0.5962	0.5509	0.4926	0.6483	0.6059	0.5506
4	75	0.6366	0.5992	0.5497	0.6765	0.6408	0.5933	0.7076	0.6739	0.6289
4	100	0.7041	0.6733	0.6319	0.7287	0.6992	0.6593	0.7492	0.7212	0.6832
5	10	0.0096	0.0033	0.0007	0.2906	0.2083	0.1308	0.5072	0.4139	0.3123
5	15	0.0763	0.0470	0.0236	0.3214	0.2518	0.1786	0.4891	0.4146	0.3290
5	20	0.1560	0.1138	0.0727	0.3618	0.2982	0.2271	0.4984	0.4333	0.3559
5	25	0.2283	0.1806	0.1296	0.4006	0.3413	0.2723	0.5151	0.4561	0.3843
5	30	0.2906	0.2412	0.1852	0.4359	0.3803	0.3136	0.5336	0.4792	0.4117
5	40	0.3895	0.3412	0.2828	0.4960	0.4463	0.3844	0.5699	0.5222	0.4615
5	50	0.4634	0.4182	0.3616	0.5443	0.4994	0.4421	0.6024	0.5596	0.5042
5	75	0.5854	0.5482	0.4994	0.6310	0.5950	0.5473	0.6662	0.6319	0.5863
5	100	0.6600	0.6289	0.5872	0.6885	0.6584	0.6180	0.7120	0.6832	0.6444

(cont'd on next page)

TABLE 3-10 (cont'd)

		df = 0			df = 10			df = 20		
k	$n\backslash\alpha$	0.05	0.01	0.001	0.05	0.01	0.001	0.05	0.01	0.001
6	15	0.0451	0.0254	0.0112	0.2790	0.2121	0.1445	0.4562	0.3798	0.2943
6	20	0.1123	0.0784	0.0472	0.3160	0.2550	0.1887	0.4602	0.3945	0.3181
6	25	0.1786	0.1375	0.0949	0.3535	0.2964	0.2313	0.4748	0.4156	0.3448
6	30	0.2382	0.1938	0.1448	0.3886	0.3346	0.2710	0.4925	0.4381	0.3715
6	40	0.3364	0.2910	0.2371	0.4495	0.4007	0.3407	0.5289	0.4811	0.4209
6	50	0.4121	0.3685	0.3146	0.4995	0.4550	0.3988	0.5624	0.5193	0.4641
6	75	0.5403	0.5033	0.4551	0.5909	0.5546	0.5070	0.6296	0.5948	0.5488
6	100	0.6205	0.5890	0.5471	0.6526	0.6220	0.5811	0.6787	0.6493	0.6099
7	15	0.0249	0.0127	0.0049	0.2463	0.1821	0.1194	0.4331	0.3543	0.2685
7	20	0.0794	0.0534	0.0304	0.2786	0.2207	0.1592	0.4295	0.3636	0.2882
7	25	0.1389	0.1045	0.0700	0.3141	0.2597	0.1989	0.4410	0.3822	0.3128
7	30	0.1949	0.1562	0.1143	0.3481	0.2966	0.2367	0.4573	0.4033	0.3381
7	40	0.2907	0.2493	0.2007	0.4088	0.3617	0.3044	0.4929	0.4455	0.3864
7	50	0.3668	0.3261	0.2761	0.4596	0.4163	0.3620	0.5266	0.4839	0.4294
7	75	0.4992	0.4635	0.4172	0.5542	0.5184	0.4716	0.5960	0.5612	0.5155
7	100	0.5838	0.5530	0.5120	0.6192	0.5888	0.5481	0.6477	0.6182	0.5786
8	20	0.0546	0.0347	0.0182	0.2474	0.1913	0.1335	0.4047	0.3373	0.2618
8	25	0.1070	0.0780	0.0498	0.2805	0.2279	0.1703	0.4125	0.3530	0.2841
8	30	0.1591	0.1247	0.0884	0.3133	0.2632	0.2060	0.4270	0.3727	0.3081
8	40	0.2517	0.2130	0.1682	0.3732	0.3270	0.2715	0.4611	0.4136	0.3550
8	50	0.3275	0.2884	0.2409	0.4243	0.3815	0.3284	0.4947	0.4518	0.3977
8	75	0.4628	0.4273	0.3817	0.5212	0.4853	0.4387	0.5656	0.5305	0.4845
8	100	0.5510	0.5199	0.4789	0.5889	0.5581	0.5172	0.6193	0.6193	0.5494
9	20	0.0365	0.0217	0.0104	0.2210	0.1666	0.1123	0.3838	0.3146	0.2391
9	25	0.0818	0.0576	0.0351	0.2513	0.2005	0.1462	0.3874	0.3274	0.2591
9	30	0.1294	0.0992	0.0682	0.2827	0.2342	0.1798	0.3999	0.3455	0.2816
9	40	0.2178	0.1820	0.1413	0.3414	0.2963	0.2428	0.4324	0.3850	0.3270
9	50	0.2924	0.2553	0.2107	0.3924	0.3503	0.2986	0.4658	0.4228	0.3691
9	75	0.4290	0.3943	0.3499	0.4909	0.4551	0.4089	0.5377	0.5023	0.4564
9	100	0.5200	0.4891	0.4485	0.5608	0.5298	0.4888	0.5931	0.5627	0.5225
10	20	0.0230	0.0126	0.0054	0.1982	0.1455	0.0947	0.3671	0.2961	0.2204
10	25	0.0611	0.0414	0.0239	0.2258	0.1769	0.1258	0.3658	0.3055	0.2379
10	30	0.1043	0.0780	0.0518	0.2557	0.2089	0.1574	0.3761	0.3218	0.2589
10	40	0.1881	0.1551	0.1181	0.3130	0.2692	0.2178	0.4068	0.3596	0.3026
10	50	0.2612	0.2260	0.1842	0.3636	0.3225	0.2723	0.4396	0.3970	0.3439
10	75	0.3986	0.3645	0.3212	0.4632	0.4277	0.3822	0.5121	0.4768	0.4311
10	100	0.4920	0.4612	0.4209	0.5349	0.5039	0.4631	0.5688	0.5384	0.4982

† If the calculated ratio is less than the appropriate ratio given in this table, the values are rejected as outliers.

Reprinted with permission. Copyright © by Gordon and Breach Science Publishers.

(i.e., a change in variance of the observations), or both. Ferguson (Refs. 15 and 16) has studied the power of the various rejection rules relative to both changes in level or scale. For several outliers and repeated rejection of observations, Ferguson points out that the sample coefficient of skewness $\sqrt{b_1}$

$$\sqrt{b_1} = \sqrt{n}\sum_{i=1}^{n}(x_i - \bar{x})^3/[(n-1)^{3/2}s^3]$$

$$= \sqrt{n}\left\{\sum_{i=1}^{n}(x_i - \bar{x})^3/[\Sigma(x_i - \bar{x})^2]^{3/2}\right\} \tag{3-56}$$

should be used for one-sided tests (change in level of several observations in the same direction). On the other hand, the sample coefficient of kurtosis b_2

$$b_2 = n\sum_{i=1}^{n}(x_i - \bar{x})^4/[(n-1)^2 s^4]$$

$$= n\sum_{i=1}^{n}(x_i - \bar{x})^4/[\Sigma(x_i - \bar{x})^2]^2 \tag{3-57}$$

is recommended for two-sided tests (change in level to higher and lower values) and also for changes in scale (variance). In applying the skewness and/or kurtosis tests, the $\sqrt{b_1}$ or the b_2, or both, are computed. If their observed values exceed those for significance levels given in either Table 3-11 or Table 3-12, the observation farthest from the mean is rejected and the same procedure is repeated until no further sample values are judged as outliers. (As we have said, and is well-known, $\sqrt{b_1}$ and b_2 are also used as tests of normality.)

In Eqs. 3-56 and 3-57 for $\sqrt{b_1}$ and b_2, respectively, s is defined as generally used in this chapter with $(n-1)$ df, i.e.,

$$s = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1). \tag{3-58}$$

The significance levels in Tables 3-11 and 3-12 for sample sizes of 5, 10, 15, and 20 (and 25 for b_2) were obtained by Ferguson (Refs. 15 and 16) on an IBM 704 computer using a sampling experiment or Monte Carlo procedure. The significance levels for the other sample sizes are from E. S. Pearson, "Table of Percentage Points of $\sqrt{b_1}$ and b_2 in Normal Samples; a Round Off" (Ref. 26). For $n = 25$, Ferguson's Monte Carlo values of b_2 agree with Pearson's computed values. Other tables of interest concerning $\sqrt{b_1}$ and b_2 are those of Mulholland (Ref. 27).

The $\sqrt{b_1}$ and b_2 statistics have the optimum property of being "locally" best against one-sided and two-sided alternatives, respectively. The $\sqrt{b_1}$ test is good for up to 50% spurious observations in the sample for the one-sided case, and the b_2 test is optimum in the two-sided alternatives case for up to 21% "contamination" of sample values. For only one or two outliers, however, the sample statistics of the previous paragraphs (pars. 3-5.1 and 3-5.4) are recommended, and, in fact, Ferguson (Ref. 1) discusses in detail their optimum properties of *pointing out* either one or two outliers.

Instead of the more complicated $\sqrt{b_1}$ and b_2 statistics, one can use the Tietjen and Moore tests discussed in par. 3-5.5.2 or Rosner's test from par. 3-5.5.3 and Hawkins' test from par. 3-5.5.3 for the sample sizes and percentage points given.

3-6 RECOMMENDED OUTLIER TESTS USING INDEPENDENT STANDARD DEVIATION ESTIMATORS

We now consider tests of outliers for which the estimate of variance is independent of the suspected values tested in samples. Such tests apply, for example, to analysis of variance tables and elsewhere. In par. 3-5.5.3 we also mentioned some related concepts by Hawkins (Ref. 21).

Suppose that an independent estimate of the standard deviation is available from either previous data or is otherwise available, as under null hypothesis situations for the analyses of variance (ANOVA's).

TABLE 3-11

SIGNIFICANCE LEVELS FOR $\sqrt{b_1}$

Significance Level, %	Sample Size n									
	5[a]	10[a]	15[a]	20[a]	25	30	35	40	50	60
1	1.34	1.31	1.20	1.11	1.06	0.98	0.92	0.87	0.79	0.72
5	1.05	0.92	0.84	0.79	0.71	0.66	0.62	0.59	0.53	0.49

[a]These values were obtained by Ferguson (Refs. 15 and 16) using a Monte Carlo procedure.

Reprinted with permission. Copyright © for portion of table by Biometrika Trustees; copyright © for remainder of table by University of California Press.

TABLE 3-12

SIGNIFICANCE LEVELS FOR b_2

Significance Level, %	Sample Size n							
	5[a]	10[a]	15[a]	20[a]	25[a]	50	75	100
1	3.11	4.83	5.09	5.23	5.00	4.88	4.59	4.39
5	2.89	3.85	4.07	4.15	4.00	3.99	3.87	3.77

[a]These values were obtained by Ferguson (Refs. 15 and 16) using a Monte Carlo procedure.

Reprinted with permission. Copyright © for portion of table by Biometrika Trustees; copyright © for remainder of table by University of California Press.

These estimates of the true σ may be from a single sample of previous similar data, or they may be the result of combining estimates from several such previous sets of appropriate data. In any event each such estimate will have df equal to one less than the sample size or group on which it is based. Thus the proper combined estimate is a weighted average of the several values of s^2; the weights are proportional to the respective df. The total df in the combined estimate then is the sum of the individual df. When one uses an independent estimate of the standard deviation s_ν based on ν df, the useful test criterion recommended for judging a low or high outlier is either

$$T_1' = \frac{\bar{x} - x_1}{s_\nu} \tag{3-59}$$

or

$$T_n' = \frac{x_n - \bar{x}}{s_\nu} \tag{3-60}$$

where

ν = total number of df in the independent estimate s_ν of σ.

The critical values for T_1' and T_n' for the 5% and 1% significance levels are from David (Ref. 28) and are given in Table 3-13. In Table 3-13 the notation ν = df indicates the total number of df associated with the independent estimate of the standard deviation σ, and n indicates the number of observations in the sample under study.

Another very useful set of tables for testing samples for outlying observations using an independent s_ν is that of Halperin, Greenhouse, Cornfield, and Zalokar (Ref. 29). They have tabulated the percentage points of the statistic d, where

TABLE 3-13

CRITICAL VALUES FOR T' WHEN STANDARD DEVIATION s_ν IS INDEPENDENT OF PRESENT SAMPLE (Ref. 28)

ν = df	n								
	3	4	5	6	7	8	9	10	12
	1% Point								
10	2.78	3.10	3.32	3.48	3.62	3.73	3.82	3.90	4.04
11	2.72	3.02	3.24	3.39	3.52	3.63	3.72	3.79	3.93
12	2.67	2.96	3.17	3.32	3.45	3.55	3.64	3.71	3.84
13	2.63	2.92	3.12	3.27	3.38	3.48	3.57	3.64	3.76
14	2.60	2.88	3.07	3.22	3.33	3.43	3.51	3.58	3.70
15	2.57	2.84	3.03	3.17	3.29	3.38	3.46	3.53	3.65
16	2.54	2.81	3.00	3.14	3.25	3.34	3.42	3.49	3.60
17	2.52	2.79	2.97	3.11	3.22	3.31	3.38	3.45	3.56
18	2.50	2.77	2.95	3.08	3.19	3.28	3.35	3.42	3.53
19	2.49	2.75	2.93	3.06	3.16	3.25	3.33	3.39	3.50
20	2.47	2.73	2.91	3.04	3.14	3.23	3.30	3.37	3.47
24	2.42	2.68	2.84	2.97	3.07	3.16	3.23	3.29	3.38
30	2.38	2.62	2.79	2.91	3.01	3.08	3.15	3.21	3.30
40	2.34	2.57	2.73	2.85	2.94	3.02	3.08	3.13	3.22
60	2.29	2.52	2.68	2.79	2.88	2.95	3.01	3.06	3.15
120	2.25	2.48	2.62	2.73	2.82	2.89	2.95	3.00	3.08
∞	2.22	2.43	2.57	2.68	2.76	2.83	2.88	2.93	3.01
	5% Points								
10	2.01	2.27	2.46	2.60	2.72	2.81	2.89	2.96	3.08
11	1.98	2.24	2.42	2.56	2.67	2.76	2.84	2.91	3.03
12	1.96	2.21	2.39	2.52	2.63	2.72	2.80	2.87	2.98
13	1.94	2.19	2.36	2.50	2.60	2.69	2.76	2.83	2.94
14	1.93	2.17	2.34	2.47	2.57	2.66	2.74	2.80	2.91
15	1.91	2.15	2.32	2.45	2.55	2.64	2.71	2.77	2.88
16	1.90	2.14	2.31	2.43	2.53	2.62	2.69	2.75	2.86
17	1.89	2.13	2.29	2.42	2.52	2.60	2.67	2.73	2.84
18	1.88	2.11	2.28	2.40	2.50	2.58	2.65	2.71	2.82
19	1.87	2.11	2.27	2.39	2.49	2.57	2.64	2.70	2.80
20	1.87	2.10	2.26	2.38	2.47	2.56	2.63	2.68	2.78
24	1.84	2.07	2.23	2.34	2.44	2.52	2.58	2.64	2.74
30	1.82	2.04	2.20	2.31	2.40	2.48	2.54	2.60	2.69
40	1.80	2.02	2.17	2.28	2.37	2.44	2.50	2.56	2.65
60	1.78	1.99	2.14	2.25	2.33	2.41	2.47	2.52	2.61
120	1.76	1.96	2.11	2.22	2.30	2.37	2.43	2.48	2.57
∞	1.74	1.94	2.08	2.18	2.27	2.33	2.39	2.44	2.52

Reprinted with permission. Copyright © by Biometrika Trustees.

$$d = \max\left[\left(\frac{x_n - \bar{x}}{s_\nu}\right), \left(\frac{\bar{x} - x_1}{s_\nu}\right)\right] \tag{3-61}$$

and the standard deviation s_ν is calculated from past or other data independent of the current sample for which outliers are being tested. The authors refer to their test as that for the studentized maximum absolute deviate in normal samples. The statistic d can be seen to be that of the two-sided alternative Student-type test of Nair (Ref. 30) or Grubbs (Ref. 9), in which the scaling statistic s_ν of the denominator must be independent of the numerator residuals.

As pointed out by Halperin, Greenhouse, Cornfield, and Zalokar (Ref. 29), their tables, reproduced here as Table 3-14, may be used to test whether the largest observation without regard to sign is too large, or the tables may be used for multiple significance tests of a set of n sample means arising from independent normal populations possibly with different true means. Thus Table 3-14 may be used in many ANOVA test procedures to determine or judge either high or low treatment effects, for example.

For each entry in Table 3-14 and for any given sample size n and number of df ν, the authors of Ref. 29 list upper and lower values, these being due to the computational procedure available (see Section 3 of Ref. 29). The authors point out that the lower values are known to be closer to the true, or correct, percentage points; accordingly, they recommend using the lower tabulated levels of significance in most cases. In fact, the actual difference in exact probabilities between the two tabulated values appears to be in the second decimal place, except for the rather small sample sizes, and consequently is of little practical interest.

The reader might note that so far in the outlier-type detection procedures of this paragraph, information in the particular sample tested for outliers is not used. Therefore, one would wonder whether there would be any gain in information or perhaps in power to detect spurious values if the variability measure for the current sample were also included in the test. In this connection, the reader perhaps noticed that just this rather useful concept was available for application in Table 3-10 prepared by Hawkins (Ref. 21) for multiple tests of outliers. Hence with reference to the studentized residuals-type tests of outliers, Hawkins and Perold (Ref. 31) have prepared a table of percentage points or critical levels of the statistic

$$B^* = \max|(x_i - \bar{x})|/S = \max\left[\left(\frac{x_n - \bar{x}}{S_H}\right), \left(\frac{\bar{x} - x_1}{S_H}\right)\right] \tag{3-62}$$

where

$$S_H^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + U = S^2 + U \tag{3-63}$$

and

$$U/\sigma^2 = \nu s_\nu^2/\sigma^2 = \chi_\nu^2 \tag{3-64}$$

Thus and as before, the quantity U is an independent $\sigma^2 x_\nu^2$ variate with ν df if such information is available for use. Note also that S^2 is the total SS for the current sample of interest, which may contain contaminated values. When only data on the current or same sample are available, U (and ν) are taken as zero.

Hawkins and Perold's critical values or percentage points of their statistic B^* are given in Table 3-15.

Summarizing somewhat at this point, we note that there are a variety of useful tests and related tables to detect outliers in samples for the case in which only an independent estimate of the underlying sigma is used or for the case in which the independent estimate is used along with the current sample information.

Now—that we have covered David's statistic (Ref. 28), using an independent estimate of the standard deviation to test for an outlier; also the similar d statistic of Halperin, Greenhouse, Cornfield, and Zalokor (Ref. 29); and finally the augmented B^* statistic of Hawkins and Perold (Ref. 31)—it would be of interest to give an illustrative example. For this purpose, we will return to the interlaboratory, or round robin, test data of Table 2-7.

TABLE 3-14

UPPER AND LOWER LIMITS FOR 1% POINT OF DISTRIBUTION OF $d = \max\left[\left(\frac{x_n - \bar{x}}{s}\right), \left(\frac{\bar{x} - x_1}{s}\right)\right]$ (Ref. 29)

$\nu \backslash n$	3	4	5	6	7	8	9	10	15	20	30	40	60
3	6.97 6.08												
4	5.10 4.65	5.85 5.27											
5	4.28 4.03	4.85 4.54	5.27 4.87										
6	3.84 3.65	4.32 4.10	4.66 4.41	4.93 4.64									
7	3.56 3.42	3.98 3.83	4.28 4.13	4.52 4.33	4.71 4.49								
8	3.37 3.26	3.76 3.64	4.03 3.91	4.24 4.10	4.42 4.25	4.56 4.38							
9	3.23 3.14	3.59 3.50	3.85 3.76	4.04 3.94	4.20 4.08	4.32 4.21	4.44 4.30						
10	3.12 3.04	3.47 3.40	3.71 3.63	3.90 3.82	4.05 3.94	4.16 4.07	4.27 4.15	4.35 4.23					
15	2.84 2.79	3.14 3.10	3.34 3.30	3.49 3.45	3.61 3.57	3.71 3.67	3.79 3.75	3.87 3.81	4.15 4.07				
20	2.72 2.68	3.00 2.97	3.18 3.16	3.32 3.29	3.43 3.40	3.51 3.49	3.59 3.56	3.65 3.62	3.89 3.85	4.05 4.00			
30	2.60 2.57	2.86 2.84	3.03 3.01	3.16 3.14	3.26 3.24	3.34 3.32	3.40 3.38	3.46 3.44	3.67 3.65	3.80 3.78	3.98 3.95		
40	2.55 2.52	2.80 2.78	2.96 2.94	3.08 3.07	3.18 3.16	3.25 3.24	3.31 3.30	3.37 3.36	3.57 3.56	3.69 3.68	3.85 3.83	3.97 3.95	
60	2.50 2.47	2.74 2.72	2.89 2.88	3.01 2.99	3.10 3.09	3.17 3.16	3.23 3.22	3.28 3.27	3.47 3.46	3.59 3.58	3.74 3.73	3.85 3.83	3.98 3.97
120	2.45 2.42	2.68 2.66	2.83 2.82	2.94 2.93	3.03 3.02	3.10 3.09	3.15 3.14	3.20 3.20	3.38 3.37	3.49 3.48	3.62 3.62	3.73 3.71	3.87 3.86
∞	2.40 2.38	2.62 2.61	2.76 2.76	2.87 2.87	2.95 2.95	3.02 3.02	3.07 3.07	3.12 3.12	3.29 3.29	3.39 3.39	3.53 3.53	3.62 3.62	3.73 3.73

(cont'd on next page)

TABLE 3-14 (cont'd)

$\nu \backslash n$	3	4	5	6	7	8	9	10	15	20	30	40	60
3	3.97 3.35												
4	3.24 2.89	3.74 3.27											
5	2.89 2.65	3.30 3.01	3.61 3.24										
6	2.68 2.50	3.05 2.84	3.32 3.06	3.53 3.22									
7	2.55 2.40	2.89 2.73	3.13 2.94	3.32 3.10	3.48 3.22								
8	2.46 2.33	2.78 2.64	3.00 2.85	3.18 3.01	3.32 3.13	3.44 3.22							
9	2.40 2.28	2.70 2.58	2.91 2.78	3.07 2.94	3.21 3.06	3.32 3.15	3.41 3.21						
10	2.34 2.24	2.63 2.53	2.84 2.73	2.99 2.88	3.12 2.99	3.23 3.08	3.32 3.15	3.40 3.23					
15	2.20 2.12	2.46 2.39	3.64 2.57	2.77 2.71	2.88 2.81	2.97 2.90	3.05 2.97	3.12 3.03	3.37 3.27				
20	2.13 2.07	2.38 2.32	2.55 2.50	2.67 2.62	2.77 2.73	2.86 2.81	2.93 2.88	2.99 2.94	3.22 3.15	3.38 3.29			
30	2.07 2.01	2.30 2.26	2.46 2.42	2.58 2.54	2.67 2.64	2.75 2.72	2.82 2.78	2.88 2.84	3.08 3.04	3.22 3.17	3.40 3.34		
40	2.04 1.99	2.27 2.23	2.42 2.39	2.54 2.50	2.63 2.60	2.70 2.68	2.77 2.73	2.82 2.79	3.02 2.99	3.15 3.11	3.32 3.27	3.43 3.39	
60	2.01 1.96	2.23 2.20	2.38 2.35	2.49 2.47	2.58 2.56	2.65 2.63	2.71 2.69	2.77 2.75	2.96 2.93	3.08 3.05	3.24 3.21	3.35 3.32	3.49 3.45
120	1.98 1.93	2.20 2.17	2.34 2.32	2.45 2.43	2.54 2.52	2.61 2.59	2.66 2.64	2.72 2.70	2.89 2.88	3.01 2.99	3.16 3.14	3.27 3.25	3.40 3.38
∞	1.95 1.91	2.16 2.14	2.30 2.28	2.41 2.39	2.49 2.48	2.56 2.55	2.61 2.60	2.66 2.65	2.83 2.82	2.95 2.94	3.09 3.08	3.19 3.18	3.31 3.30

Reprinted with permission. Copyright © by the American Statistical Association.

TABLE 3-15

PERCENTAGE POINTS OF HAWKINS' B^* (Ref. 31)

	$\nu = 0$				$\nu = 5$				$\nu = 15$				$\nu = 30$			
α	0.10	0.05	0.01	0.001	0.10	0.05	0.01	0.001	0.10	0.05	0.01	0.001	0.10	0.05	0.01	0.001
n																
5	0.8357	0.8575	0.8818	0.8917	0.6396	0.6839	0.7573	0.8190	0.4583	0.5011	0.5796	0.6606	0.3468	0.3827	0.4510	0.5264
6	0.8149	0.8440	0.8823	0.9032	0.6380	0.6813	0.7554	0.8207	0.4690	0.5103	0.5869	0.6674	0.3593	0.3942	0.4612	0.5360
7	0.7912	0.8246	0.8733	0.9051	0.6325	0.6750	0.7493	0.8173	0.4749	0.5148	0.5898	0.6696	0.3678	0.4018	0.4676	0.5417
8	0.7679	0.8038	0.8596	0.9006	0.6251	0.6667	0.7409	0.8109	0.4777	0.5165	0.5901	0.6690	0.3736	0.4069	0.4715	0.5448
9	0.7438	0.7831	0.8439	0.8923	0.6163	0.6576	0.7314	0.8029	0.4787	0.5165	0.5886	0.6668	0.3777	0.4102	0.4738	0.5463
10	0.7254	0.7633	0.8274	0.8817	0.6081	0.6480	0.7213	0.7939	0.4784	0.5152	0.5861	0.6635	0.3805	0.4124	0.4750	0.5466
11	0.7064	0.7445	0.8108	0.8697	0.5993	0.6385	0.7111	0.7845	0.4771	0.5131	0.5828	0.5595	0.3824	0.4137	0.4753	0.5461
12	0.6889	0.7271	0.7947	0.8571	0.5906	0.6290	0.7010	0.7748	0.4752	0.5105	0.5790	0.6550	0.3836	0.4143	0.4750	0.5451
13	0.6728	0.7107	0.7791	0.8442	0.5820	0.6197	0.6910	0.7651	0.4729	0.5075	0.5749	0.6501	0.3842	0.4144	0.4742	0.5436
14	0.6578	0.6954	0.7642	0.8313	0.5737	0.6107	0.6812	0.7554	0.4703	0.5042	0.5706	0.6451	0.3843	0.4140	0.4731	0.5417
15	0.6438	0.6811	0.7500	0.8186	0.5656	0.6020	0.6717	0.7459	0.4674	0.5007	0.5662	0.6400	0.3842	0.4134	0.4717	0.5397
16	0.6308	0.6676	0.7364	0.8062	0.5578	0.5936	0.6625	0.7355	0.4544	0.4971	0.5617	0.6348	0.3837	0.4125	0.4701	0.5374
17	0.6187	0.6550	0.7235	0.7942	0.5503	0.5855	0.6535	0.7274	0.4612	0.4935	0.5571	0.6295	0.3831	0.4115	0.4683	0.5350
18	0.6073	0.6431	0.7112	0.7825	0.5430	0.5777	0.6449	0.7185	0.4580	0.4898	0.5526	0.6243	0.3822	0.4103	0.4665	0.5325
19	0.5965	0.6319	0.6996	0.7711	0.5351	0.5701	0.6366	0.7099	0.4548	0.4860	0.5480	0.6192	0.3812	0.4089	0.4645	0.5299
20	0.5864	0.6213	0.6884	0.7602	0.5293	0.5629	0.6285	0.7015	0.4516	0.4823	0.5436	0.6141	0.3801	0.4075	0.4624	0.5272
21	0.5769	0.6113	0.6778	0.7497	0.5229	0.5559	0.6209	0.6934	0.4483	0.4786	0.5392	0.6090	0.3789	0.4059	0.4603	0.5245
22	0.5679	0.6018	0.6677	0.7396	0.5166	0.5492	0.6134	0.6855	0.4451	0.4750	0.5348	0.6040	0.3776	0.4043	0.4581	0.5218
23	0.5593	0.5927	0.6581	0.7298	0.5106	0.5427	0.6062	0.6778	0.4419	0.4714	0.5305	0.5991	0.3763	0.4027	0.4559	0.5191
24	0.5512	0.5841	0.6488	0.7204	0.5048	0.5365	0.5993	0.6704	0.4387	0.4678	0.5263	0.5943	0.3749	0.4010	0.4537	0.5163
25	0.5434	0.5750	0.6400	0.7113	0.4992	0.5305	0.5925	0.6632	0.4356	0.4643	0.5221	0.5896	0.3735	0.3993	0.4515	0.5136
26	0.5350	0.5681	0.6315	0.7025	0.4938	0.5247	0.5861	0.5552	0.4325	0.4609	0.5180	0.5850	0.3720	0.3975	0.4492	0.5109
27	0.5289	0.5607	0.6234	0.5940	0.4886	0.5190	0.5798	0.6495	0.4294	0.4575	0.5140	0.5804	0.3705	0.3958	0.4470	0.5081
28	0.5222	0.5535	0.6156	0.6859	0.4836	0.5136	0.5737	0.6429	0.4264	0.4541	0.5101	0.5760	0.3690	0.3940	0.4447	0.5054
29	0.5157	0.5466	0.6081	0.6780	0.4787	0.5084	0.5678	0.6365	0.4234	0.4509	0.5063	0.5716	0.3675	0.3922	0.4425	0.5027
30	0.5095	0.5400	0.6009	0.6704	0.4740	0.5033	0.5621	0.5303	0.4205	0.4476	0.5025	0.5673	0.3660	0.3904	0.4403	0.5001

Reprinted with permission. Copyright © by University of Manitoba.

Example 3-11:

In the interlaboratory test of par. 2-10 for measurements of the amount of lead in gasoline, it seemed probable that the levels of measurement of the Du Pont and Mobil laboratories were low compared to those of the other laboratories. Is there any statistical evidence to back up this hypothesis?

Since, under the assumptions of the ANOVA procedure, the among-laboratory and within-laboratory SS are independent, we will first use only the residual or within-laboratory SS to estimate sigma. In this connection, we have that $\sigma_r = 0.50$ based on the within-laboratory SS of 2.50 and $\nu = 10$ df.

The observed levels or average measurements of the amount of lead in gasoline (multiplied by 1000) are as follows:

Du Pont	Mobil	EPA	Ethyl	Ford	AMOCO	Octel
23.3	24.0	25.7	26.0	26.7	27.5	28.0.

We note, however, that these were based on different sample sizes, i.e., either 2 or 3 per laboratory. A very satisfactory, approximate way to solve the problem posed is to note that the grand mean for all the laboratories is $\bar{\bar{x}} = 438/17 = 25.76$; therefore, we will consider the largest deviations from this value. In fact, we may as well pool the readings of Du Pont and Mobil since we will test both as low outliers and obtain their average as

$$(70 + 48)/5 = 23.60.$$

Hence we will use an approximate test on the difference

$$25.76 - 23.60 = 2.16$$

and we must determine the estimated standard error of this unevenly weighted difference. Under the null hypothesis of no differences in laboratory levels and hence the use of only the within-laboratory sigma for testing for outliers, we note that the stated difference is really

$$118/5 - [(118) + (438 - 118)]/17 = \left(\frac{1}{5} - \frac{1}{17}\right)(118) - \frac{1}{17}(320) = \frac{12}{85}(118) - \frac{1}{17}(320) = -2.16$$

where 118 is the sum of 5 observations of Du Pont and Mobil, and 320 is the sum of the remaining 12 observations. Thus since σ_r^2 is the variance of an individual laboratory reading, the estimated variance of the stated difference, i.e., -2.16, is

$$\sigma^2(\text{diff}) = \left(\frac{12}{85}\right)^2(5\sigma_r^2) + \left(\frac{5}{85}\right)^2(12\sigma_r^2) = 0.141\sigma_r^2.$$

This means that the equivalent sample size for the numerator of a Student's t-type statistic to use is about $1/0.141 = 7.09$. Hence we may take our studentized statistic t to be approximately

$$t \approx -2.16/(\sqrt{0.141}\,\hat{\sigma}_r) = -2.16/(0.5/\sqrt{7.09}) = -11.50$$

which for $\nu = 10$ df is very highly significant from either Table 3-13 or Table 3-14. There seems to be little doubt, therefore, on the basis of the ANOVA residual or error variance, that the readings of Du Pont and Mobil are significantly low. The ANOVA of Table 2-7 established a very significant difference between the among-laboratory and within-laboratory variations, i.e., a huge ratio of $7.093/0.25 = 28.37$ to 1 on the variance scale or 5.33 to 1 on the sigma scale.

Ordinarily, Hawkins' B^* test might be applied to testing whether the Du Pont and Mobil laboratory levels are low if we could pool the among-laboratory and within-laboratory sum of squares. We can at least illustrate the principle in spite of the fact that there is a large difference between the among- and within-laboratory variances. Thus we found the sum of squares (about the table mean) among columns based on an individual reading to be 42.56 and that of the within or residual sum of squares U to be 2.50. Hence according to Eq. 3-63, we obtain

$$S^2 = 42.56 + 2.50 = 45.06$$

where the ν of Eq. 3-64 has the value, $\nu = 10$ df. Since the average 23.60 was based on the *equivalent* of about 7.09 observations and $S^2 = 45.06$ is for an individual observation, we take Hawkins' B^* as approximately $B^* \approx \sqrt{7.09}\,(\bar{\bar{x}} - \bar{x}_1)/S = 2.66\ (25.76 - 23.60)\ /6.71 = 0.86$, where we used the grand mean $\bar{\bar{x}}$ and the average of the two lowest laboratories. Referring to Table 3-15 for critical values of Hawkins' B^*, we find for $n = 6$ laboratories (we combined Du Pont and Mobil) and $\nu = 5$ df that the 0.001 percentage point is 0.8207, whereas for $\nu = 15$, the 0.001 probability level is 0.6674. Therefore, for $\nu = 10$ we would even reject the null hypothesis of no difference among laboratory measurements under the (questionable) pooling procedure. In any event, it certainly seems that we can now settle the question raised in Table 2-7; namely, the measurements of lead in gasoline by Du Pont and Mobil are significantly low, and an investigation is called for to "bring them into line". (All laboratories, on the average, still measure a little low.)

It is such an investigation of laboratory measurement levels that is called for concerning the whole matter of testing for outlying laboratories. Thus we saw in Table 2-7 that the within, residual, or repeatability sigma amounted to 0.50 and the among-laboratory sigma had a value of 1.69, so that the reproducibility sigma for an individual measurement taken at a randomly selected laboratory became 1.76. This shows that the residual sigma representing precision at one or a single laboratory is quite inconsequential because practically all the variability comes from the fact that the laboratory levels are not in agreement, and, therefore, there is indeed quite a problem to bring them together or to calibrate their measurement procedures or instruments. This is at the heart of the whole matter of procedures for testing for aberrant readings, and we see that it becomes urgent to investigate first and to do something about the results coming from Du Pont and Mobil. In fact, it is only through such investigations or through calibration procedures that we can hope to reduce the among-laboratory sigma of 1.69 and thereby gain some improvement in the precision of measurement of the amount of lead in gasoline.

In addition, it is easy to note that although we had no problem really in the choice of the "right" underlying estimate of sigma to test for outliers in single samples, this is not the case for ANOVA procedures where two or more components of variance may be real and quite different, as in Table 2-7. In fact, we believe that the among-laboratory sigma may not be brought into line with the almost negligible residual sigma of only 0.50. That is, we should expect that the among-laboratory sigma will most always be larger than the within value at a single laboratory, and, in fact, several times the latter value. Hence we should expect that this would be the usual case and that the real or basic problem toward improving precision and accuracy would revolve around properly correcting for the different measurement levels at the various laboratories. Having observed this, we will proceed with another, but more extensive, example (Example 3-12) on interlaboratory testing and will show that our thoughts on the matter are well verified and justified.

Example 3-12:

In an analysis of interlaboratory test procedures, data representing normalities of sodium hydroxide solutions were determined by 12 different laboratories. In all the standardizations a 0.1 normal sodium hydroxide solution was prepared by the Standard Methods Committee using carbon-dioxide-free distilled water. Potassium acid phthalate (PAP), obtained from the National Bureau of Standards, was used as the test standard at all of the participating laboratories in the round robin test.

Test data by the 12 laboratories are given in Table 3-16. The PAP readings have been coded to simplify the calculations. The variances among the three readings within all laboratories were found to be homogeneous. A one-way classification in the ANOVA was first analyzed to determine whether the variation in laboratory results (averages) was statistically significant. This variation was found to be very significant and indicated a need for action, so tests for outliers were then applied to isolate the particular laboratories whose results gave rise to the significant variation.

Table 3-17 shows that the variation between laboratories is highly significant, exhibiting an F ratio of 48.61. To test whether this (very significant) variation is caused by one laboratory (or perhaps two) that obtained "outlying" results (i.e., perhaps showing nonstandard technique), we can test the laboratory averages for outliers. From the ANOVA we have an estimate of the within or residual variance of an indi-

TABLE 3-16
STANDARDIZATION OF SODIUM HYDROXIDE SOLUTIONS AS DETERMINED BY PLANT LABORATORIES (Ref. 10)

Standard Used: Potassium Acid Phthalate (PAP)

Laboratory	(PAP 0.096000) $\times 10^3$	Sums	Averages	Deviation of Average from Grand Average
1	1.893			
	1.972			
	1.876	5.741	1.914	+0.043
2	2.046			
	1.851			
	1.949	5.846	1.949	+0.078
3	1.874			
	1.792			
	1.829	5.495	1.832	−0.039
4	1.861			
	1.998			
	1.983	5.842	1.947	+0.076
5	1.922			
	1.881			
	1.850	5.653	1.884	+0.013
6	2.082			
	1.958			
	2.029	6.069	2.023	+0.152
7	1.992			
	1.980			
	2.066	6.038	2.013	+0.142
8	2.050			
	2.181			
	1.903	6.134	2.045	+0.174
9	1.831			
	1.883			
	1.855	5.569	1.856	−0.015
10	0.735			
	0.722			
	0.777	2.234	0.745	−1.126
11	2.064			
	1.794			
	1.891	5.749	1.916	+0.045
12	2.475			
	2.403			
	2.102	6.980	2.327	+0.456
Grand Sum		67.350		
Grand Average			1.871	

Reprinted with permission. Copyright © by the American Statistical Association.

vidual reading as 0.008793 based on 24 df. The estimated standard deviation of the average of three readings is therefore $0.094/\sqrt{3} = 0.054$. The complete ANOVA is given in Table 3-17 and, due to the huge variation resulting from some differences in levels of measurement for some of the laboratories, we must now conduct an analysis to determine just which laboratories have unacceptable levels of measurement.

In this example we are not concerned about any variation in number of observations per laboratory since they are all three in number, and hence no adjustment for the 50% variation from two to three observations is needed as in Example 3-11. Also since we illustrated the Hawkins technique in Example 3-11, we may as well use David's studentized statistic or the d statistic of Halperin, Greenhouse, Cornfield, and Zalokar, and accompanying tables of percentage points. Since the estimate of within-laboratory variation is independent of any difference between laboratories, we can use the David statistic T_1' of Eq. 3-59 and T_n' of Eq. 3-60 to test for outliers. An examination of the deviations of the laboratory averages from the grand average indicates that Laboratory 10 obtained an average reading much lower than the grand average and that Laboratory 12 obtained a rather high average level of measurement compared to the overall average. First, to test whether Laboratory 10 is an outlier, we calculate

$$T_1' = \frac{1.871 - 0.745}{0.054} = 20.9.$$

The value of T_1' is, from Table 3-13, obviously significant at a very low level of probability ($P << 0.01$). We conclude, therefore, that the test methods of Laboratory 10 should be investigated and corrected.

Excluding Laboratory 10 and at the risk of increasing the Type I error*, we compute a new grand average of 1.973 and test whether the results of Laboratory 12 are outlying. We have that

$$T_n' = \frac{2.327 - 1.973}{0.054} = 6.56$$

and this value of T_n' is significant at $P << 0.01$. We conclude that the procedures of Laboratory 12 should also be investigated.

Concerning Laboratories 10 and 12, we could also have used Table 3-14 or, that is, the maximum independently studentized statistic d of Halperin, Greenhouse, Cornfield, and Zalokar (Ref. 29). In this connection we see that for Laboratory 10, $d = T_1' = 20.9$, and using Table 3-14 for $n = 12$ and $\nu = 24$, it is quite clear that Laboratory 10 is an outlier. Moreover, repeating this same test after eliminating Laboratory 10, we see also that Laboratory 12 has too high a level of measurement and should be investigated. In summary, we find that the d statistic establishes that Laboratories 10 and 12 are outliers and should be investigated. Furthermore, Halperin, Greenhouse, Cornfield, and Zalokar (Ref. 29) point out in their appendix that the chance that the statement made concerning Laboratories 10 and 12 is incorrect when the null hypothesis of no differences whatever is true is clearly 0.01—our specified level of testing. Also when the null hypothesis is false, this chance is less than 0.01, even for multiple tests.

To verify that the remaining laboratories did indeed obtain homogeneous results, we might repeat the analysis of variance omitting Laboratories 10 and 12. The calculations give the results shown in Table 3-18.

For this analysis, the variation between laboratories is not significant at the 5% level, and we conclude that all except Laboratories 10 and 12 exhibit the same capability in testing procedure.

In conclusion, there should be a systematic investigation of test methods for Laboratories 10 and 12 to determine why their test procedures are apparently different from the other ten laboratories.

*Determination and control of the Type I error, especially with the aid of the Bonferroni inequalities, is discussed in Chapter 4.

TABLE 3-17
ANALYSIS OF VARIANCE FOR THE DATA OF TABLE 3-16

Source of Variation	Degrees of Freedom (df)	Sum of Squares (SS)	Mean Square (MS)	F Ratio
Between laboratories	11	4.70180	0.4274	$F = 48.61$ (highly significant) $P < 0.001$
Within laboratories	24	0.21103	0.008793	
Total	35	4.91283		

Reprinted with permission. Copyright © by the American Statistical Association.

TABLE 3-18
ANALYSIS OF VARIANCE OMITTING LABORATORIES 10 AND 12

Source of Variation	Degrees of Freedom (df)	Sum of Squares (SS)	Mean Square (MS)	F Ratio
Between laboratories	9	0.13889	0.01543	$F = 2.35$ (not significant) $F_{0.05}(9,20) = 2.40$ $F_{0.01}(9,20) = 3.45$
Within laboratories	20	0.13107	0.00655	
Total	29	0.26996		

Reprinted with permission. Copyright © by the American Statistical Association.

3-7 RECOMMENDED CRITERIA FOR KNOWN STANDARD DEVIATION

Frequently, the population standard deviation σ may be known with sufficient accuracy and hence does not have to be estimated.

In such cases a statistic of the form

$$T'_{1\infty} = (\bar{x} - x_1)/\sigma \tag{3-65}$$

or

$$T'_{n\infty} = (x_n - \bar{x})/\sigma \tag{3-66}$$

where

$$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n$$

may be used to test for simple outliers. Table 3-19 gives the critical values of $T'_{1\infty}$ and $T'_{n\infty}$. We illustrate this with Example 3-13.

Example 3-13 (σ known):

In the early days of satellites, the passage of the Echo 1 (Balloon) Satellite was recorded on star plates when it was visible. Photographs were made by means of a camera with the shutter automatically timed to obtain a series of points for the Echo path. Since the stars were also photographed at the same times as the Satellite, all the pictures showed star trails and were thus called star plates.

The x- and y-coordinates of each point on the Echo path were read from a photograph with a stereocomparator. To eliminate bias of the reader, the photograph was placed in one position and the coordinates were read; then the photograph was rotated 180 deg and the coordinates reread. The average of the two readings was taken as the final reading. Before any further calculations were made, the readings had to be screened for gross reading or tabulation errors. This was done by examining the difference in the readings taken at the two positions of the photograph.

TABLE 3-19

CRITICAL VALUES OF $T'_{l\infty}$ AND $T'_{n\infty}$ WHEN THE POPULATION STANDARD DEVIATION σ IS KNOWN (Ref. 10)

Number of Observations n	5% Significance Level	1% Significance Level	0.5% Significance Level
2	1.39	1.82	1.99
3	1.74	2.22	2.40
4	1.94	2.43	2.62
5	2.08	2.57	2.76
6	2.18	2.68	2.87
7	2.27	2.76	2.95
8	2.33	2.83	3.02
9	2.39	2.88	3.07
10	2.44	2.93	3.12
11	2.48	2.97	3.16
12	2.52	3.01	3.20
13	2.56	3.04	3.23
14	2.59	3.07	3.26
15	2.62	3.10	3.29
16	2.64	3.12	3.31
17	2.67	3.15	3.33
18	2.69	3.17	3.36
19	2.71	3.19	3.38
20	2.73	3.21	3.39
21	2.75	3.22	3.41
22	2.77	3.24	3.42
23	2.78	3.26	3.44
24	2.80	3.27	3.45
25	2.81	3.28	3.46

Reprinted with permission. Copyright © by the American Statistical Association.

Table 3-20 records a sample of six readings made by the Ballistic Research Laboratories (BRL) at the two positions and the differences in these readings. On the third reading the differences are rather large. Has the operator made an error in placing the cross hair on the point?

For this example an independent estimate of σ is available since extensive tests on the stereocomparator have shown that the standard deviation in reader's error is about 4 μm. The standard deviation of the difference in two readings is therefore

$$\sqrt{4^2 + 4^2} = \sqrt{32} = 5.7\ \mu\text{m}.$$

For the six readings (Table 3-20) the mean difference in the x-coordinates is $\Delta x = 3.5$, and the mean difference in the y-coordinates is $\Delta y = 1.8$. By using Eq. 3-66 for the questionable third reading, we have

$$T'_x = \frac{24 - 3.5}{5.7} = 3.60$$

$$T'_y = \frac{22 - 1.8}{5.7} = 3.54.$$

TABLE 3-20

STAR PLATE MEASUREMENTS, μm*

x-coordinate			y-coordinate		
Position 1	Position 1 + 180 deg	Δx	Position 1	Position 1 + 180 deg	Δy
−53011	−53004	−7	70263	70258	+5
−38112	−38103	−9	−39729	−39723	−6
−2804	−2828	+24	81162	81140	+22
18473	18467	+6	41477	41485	−8
25507	25497	+10	1082	1076	+6
87736	87739	−3	−7442	−7434	−8

*These data represent a sample of typical measurements taken by the former Ballistic Measurements Laboratory of the BRL many years ago.

Reprinted with permission. Copyright © by the American Statistical Association.

From Table 3-19 we see that for $n = 6$ values of $T'_{1\infty}$ as large as the calculated values would occur by chance less than 1% of the time (actually even less than 0.5%) so that a significant reading error seems to have been made on x- and y-coordinate readings for the third point.

A great number of points are read and automatically tabulated on star plates. Here we have chosen a very small sample of these points. In actual practice the tabulations would probably be scanned quickly for very large errors, such as tabulator errors; then some rule-of-thumb, such as +3 standard deviations of reader's error, might be used to scan for outliers caused by operator error. (Note that the values of Table 3-19 vary between about 1.40σ and 3.50σ.) In other words, the data are probably too extensive to allow repeated use of precise tests, such as those described heretofore in this chapter (especially for varying sample size), but this example does illustrate the case where σ is known with sufficient accuracy from past information. Therefore, if gross disagreement is found in the two readings of a coordinate, the reading could be omitted or reread before further computations are made.

The tracking data analysis-type problem we have just discussed brings up a whole new area of testing, recording data, analyzing information, and investigating implications because data become very numerous indeed and lead to formidable volumes of observations to treat or process. In fact, with such large amounts of information there is hardly time to detect and search for the actual causes of aberrant observations, i.e., their physical cause, and such irregularities occur frequently. Thus the prime or pressing object in such applications may be that of developing a suitable measure of central tendency, and consequently, there might be many smoothing procedures that could be satisfactorily applied in addition to least squares discussed in Chapter 6. For small samples and especially in research and development, many investigators do not like to discard any data at all, so that one of our prime purposes in this chapter has been to indicate just when the scientist or engineer should probably stop and look for causes of aberrant sample values. However, for the tracking data analysis-type problem or for cases in which the investigator really has no real or deep interest in detecting outliers, he may well consider other methods of estimation. As a matter of fact, there is now such a proliferation of computers that many investigators may even program almost any analytical techniques they desire irrespective of any statistical or mathematical complications. In addition, there is always the concern on the part of the statistician and others about the usual or required assumption of normality. In recent years, there has been very wide interest and much statistical research on robust estimation procedures, and the interested reader might study these new areas for possible application of other statistical techniques. He might, for example, first examine the survey articles by

Huber (Ref. 32) and Hogg (Ref. 33) to acquire interest in that direction. Thus we are cautioning the investigator or applied statistician that as a result of much statistical research and the various accomplishments during the past ten years or so, there now exist many suitable procedures for the analysis of experimental data; accordingly, one may have to compare possibly applicable techniques on a rather extensive basis to determine the best methods of analysis for his particular problem.

3-8 THE WILK-SHAPIRO STATISTICAL TEST FOR NONNORMALITY

Earlier in the chapter we remarked about the somewhat close relation between tests for outliers on the one hand versus tests for normality on the other for the data presented to us for analysis. In this connection, therefore, we should include in our discussion something concerning an appropriate test for normality. Of course, there exist many, many different statistical tests for determining whether the information available in our sample of interest does indeed get a go-ahead insofar as normality is concerned. However, it is not our purpose to delve very extensively into tests or procedures for detecting departures from the assumption of normality. We will nevertheless include one of the procedures that has been found to be quite useful and sensitive toward detecting trends away from normality—i.e., the Wilk-Shapiro test (Refs. 34, 35, 36). Thus a sample test criterion for nonnormality, and hence possibly for outliers, not covered previously is the Wilk-Shapiro W statistic for a sample of size n given by

$$W = \left[\sum_{i=1}^{[n/2]} a_{n-i+1}(x_{n-i+1} - x_i)\right]^2 \Big/ \sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{3-67}$$

where

$$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n$$

$$\bar{x} = \sum_{i=1}^{n} x_i/n$$

$$[n/2] = \text{the greatest integer in } n/2.$$

The coefficients, a_{n-i+1}, of the order statistics for $n = 2(1)50$ are given in Ref. 34 as is a table of percentage points of the statistic W for $n = 3(1)50$.*

The Wilk-Shapiro W statistic has been found to be quite sensitive to departures from normality and may compare most favorably with the $\sqrt{b_1}$ and b_2 tests discussed in par. 3-5.5.4. In addition, therefore, the W statistic may be used also as a test for outliers or otherwise general heterogeneity of sample values. The significance tests given here have been selected and recommended because they generally *point out* particular suspected outliers in the sample, so that perhaps worthwhile investigations may be pursued to find causes. Indeed, it is through such investigations that progress is made in research and development. Hence we have recorded the Wilk-Shapiro test to indicate further avenues of approach to the problems of outliers and nonnormality.

3-9 PROBABILITY PLOTS AND GRAPHICAL TECHNIQUES

With the advent of the high-speed digital computer and peripheral plotting equipment—along with the generation of huge amounts of experimental- or simulation-type data in so many fields of endeavor—there has been an increasing amount of applied interest in probability plots of all kinds. For example, the sample data may be plotted on normal probability papers to determine whether the data perhaps exhibit the possible existence of a normal universe, thereby meeting this assumption. There is also probability paper or graphs to determine whether reliability or life testing-type data follow a Weibull distribution or an exponential distribution, and graphical means incorporated therewith even to estimate the population parameters of the larger category sampled. Hence, a quick and often very suitable type of statistical analysis can be made by means of using probability paper plots of data.

* Another pertinent reference, probably more readily available, is the book by G. J. Hahn and S. S. Shapiro, *Statistical Models in Engineering*, John Wiley & Sons, Inc., New York, NY, 1967 (pp. 294-302).

Daniel (Ref. 37) has used half-normal probability plots for interpreting two-level factorial experiments. Chernoff and Lieberman (Ref. 38), for example, discuss the uses of normal probability paper, and they give an account of the uses of generalized probability paper for continuous distributions in Ref. 39. D. M. Sparks (Ref. 40) discusses an account of half-normal plotting and also gives the printed computer program required for use. Zahn (Ref. 41) indicates that some modifications and revisions of percentage points or critical values used in connection with half-normal plots are required. Wilk, Gnanadesikan, and Huyett (Ref. 42) cover a discussion of probability plots for the gamma distribution. Thus these remarks should at least indicate that a growing and useful area of applications for probability plotting does indeed exist, and the reader might well study these techniques for his own applications. In fact, and in addition to other uses, it becomes clear that probability plots may be used also for detecting possible outliers in samples since any departures from the hypothesized lines on probability papers would indicate that the assumptions are probably violated. Also it is easy to see that large individual deviations might well point to outliers. We therefore suggest that interested readers might well consider the use of probability plots to detect outliers or otherwise abnormal conditions since "a picture is worth a thousand words" also in this area of investigation or analysis.

A general discussion of probability plotting methods for the analysis of data is presented by Wilk and Gnanadesikan (Ref. 43).

Along with the use of probability plots, we should mention also graphical methods or plots in connection with outlier examinations. Prescott (Ref. 44), reporting at the 1977 Sheffield (England) Conference on Graphical Methods in Statistics, presented some results on graphical examinations concerning the behavior of outlier tests when more than a single outlier is present. Prescott's graphs show rather strikingly the effect of masking, which we discussed in par. 3-2.3, along with the basic work of Pearson and Chandra Sekar (Ref. 3).

3-10 ADDITIONAL COMMENTS AND GUIDELINES

With this introductory account of the problem concerning statistical tests of significance for detecting outlying observations, the reader will likely want to extend his knowledge of the general subject matter and perhaps delve more fully into all aspects of this important topic. In fact, the detection and proper treatment of outliers or aberrant values in samples probably represents one of the central problems of statistics. Outliers cannot be ignored since in many cases they have a decided effect on inferences from the sample data. Moreover, once we have detected outliers, some action should be taken to locate causes. Corrections for these anomalous observations should follow in order that we acquire a set of data that truly represents the process or physical situation we are studying. Although investigators generally do not like to reject any observations, sometimes it may become necessary. In fact, the use of "trimmed" means, variances, etc., may lead to robustness of estimation in any further data processing. A discussion and treatment of trimmed means and outer means, and their variances is available in a paper by Prescott and Hogg (Ref. 45).

Anscombe (Ref. 46) discusses the problem and treatment of outlying observations from a different point of view than that presented in this chapter; the reader may also have some interest in his "insurance-type risk" ideas.

Many investigators will want to give less weight to outliers than the other sample values, and others would like to, and actually do, conduct additional experiments to replace aberrant observations. Also there is the school of thought that outliers should be "Winsorized" or replaced with the sample values closest to them. Others may want to use the sample median instead of the sample mean, and so on. Concerning the treatment of outliers, we also want to point out that order statistics are treated in Chapter 7 and represent a subject area of allied interest, especially in view of the fact that sample values may be truncated or censored from analysis. In this connection, see also Chapter 21 of the *Army Weapon Systems Analysis Handbook, Part One*, DARCOM-P 706-101.

The principles of least squares for Army investigators are covered in Chapter 6 of this handbook. The detection and treatment of outliers in regression studies, especially including an analysis of residuals from the fitted line or curve, represent another area for processing data containing anomalous sample values.

This topic will be discussed in Chapter 6. In this connection, Elashoff (Ref. 47) presents a study of a model for quadratic outliers in linear regression. In other words, the Elashoff (Ref. 47) paper covers situations in which the data appear to veer off above or below the fitted regression line and some further special analysis seems necessary compared to the discussion of this chapter.

Ellenberg (Ref. 48), in a study of the joint distribution of the standardized least squares residuals from a general linear regression relation, also gives some criteria for tests of outliers in the multiparameter linear least squares-type of fit, and hence his tests may be of interest in various Army applications.

Finally, we draw the reader's attention to some recent work by Green (Ref. 49) on outlier-prone and outlier-resistant types of distributions. In his interesting paper Green (Ref. 49) indicates that the normal distribution, for example, is "absolutely outlier resistant". Some distributions, such as the Poisson distribution, are relatively outlier resistant but are neither absolutely outlier resistant nor absolutely outlier prone. A distribution that is absolutely oulier prone and relatively outlier resistant is the gamma distribution. The Cauchy distribution is branded as being absolutely outlier prone and one that cannot be relatively outlier resistant.

A new book on outliers is that of Barnett and Lewis (Ref. 50).

3-11 SUMMARY

In this chapter we have introduced the Army investigator or analyst to many procedures and techniques relative to the problem of examining samples for outlying observations. The topics covered include tests for detecting single aberrant sample observations, the possibility of two outliers on either the high or low side of the sample, the situation in which the highest and lowest sample values may be different from other sample values, and finally the use of detection procedures for any number of outliers. Thus the user of this chapter has readily available many tests of significance to apply to almost any problem he faces concerning outlying observations in his daily experimental work.

Many examples have been given to illustrate the applications of the theory or methodology, and the accompanying tables of critical values for the sample statistics recommended and included for general use.

REFERENCES

1. G. W. Thomson, "Bounds for the Ratio of Range to Standard Deviation", Biometrika **42**, 268-9 (June 1955).
2. Paul A. Samuelson, "How Deviant Can You Be?". Journal of the American Statistical Association **63**, 1522-5 (December 1968).
3. E. S. Pearson and C. Chandra Sekar, "The Efficiency of Statistical Tools and a Criterion for the Rejection of Outlying Observations", Biometrika **28**, 308-20 (1936).
4. W. R. Thompson, "On a Criterion for the Rejection of Observations and the Distribution of the Ratio of the Deviation to the Sample Standard Deviation", Annals of Mathematical Statistics **6**, 214-9 (1935).
5. *Tables of the Bivariate Normal Distribution Function and Related Functions*, US Department of Commerce, National Bureau of Standards, Applied Mathematics Series 50, June 1959.
6. E. S. Pearson and H. O. Hartley, "The Probability Integral of the Range in Samples of *n* Observations from a Normal Population", Biometrika **32**, 301-10 (April 1942).
7. Julius Lieblein, "Properties of Certain Statistics Involving the Closest Pair in a Sample of Three Observations", Journal of Research of the National Bureau of Standards **48**, Research Paper 2311, 255-68 (March 1952).
8. Frank J. Anscombe and Bruce A. Barron, "Treatment of Outliers in Samples of Size Three", Journal of Research of the National Bureau of Standards-B, Mathematics and Mathematical Physics **70B**, 141-7 (April-June 1966).
9. F. E. Grubbs, "Sample Criteria for Testing Outlying Observations", Annals of Mathematical Statistics **21**, 27-58 (March 1950).
10. Frank E. Grubbs, "Procedures for Detecting Outlying Observations in Samples", Technometrics **11**, 1-21 (February 1969).
11. Frank E. Grubbs, *Procedures for Detecting Outlying Observations in Samples*, US Army Ballistic Research Laboratories Report No. 1713, April 1974.

12. *Standard Recommended Practice for Dealing with Outlying Observations*, American Society for Testing and Materials (ASTM), 1916 Race St., Philadelphia, PA 19103, 1975. (ASTM Designation E178-75, American National Standard 21.14).
13. Frank E. Grubbs and Glen Beck, "Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations", Technometrics **14**, 847-54 (November 1972).
14. W. J. Dixon, "Processing Data for Outliers", Biometrics **9**, 74-89 (March 1953).
15. Thomas S. Ferguson, "On the Rejection of Outliers", *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, edited by Jerzy Neyman. University of California Press, Berkeley and Los Angeles, 253-87 (1961).
16. Thomas S. Ferguson,"Rules for Rejection of Outliers", Revue Institute Internationale de Statistique **29**, 29-43 (1961).
17. H. A. David, H. O. Hartley, and E. S. Pearson, "The Distribution of the Ratio in a Single Sample of Range to Standard Deviation", Biometrika **41**, 482-93 (1954).
18. G. L. Tietjen and R. H. Moore, "Some Grubbs-Type Statistics for the Detection of Several Outliers", Technometrics **14**, 583-97 (August 1972).
19. A. Kudo, "On the Testing of Outlying Observations", Sankhyā, The Indian Journal of Statistics **17**, Part 1, 67-76 (June 1956).
20. Bernard Rosner, "On the Detection of Many Outliers", Technometrics **17**, 221-7 (May 1975).
21. Douglas M. Hawkins, *Identification of Outliers*, Chapman and Hall, Ltd., 11 New Fetter Lane, London EC4P4EE, England, 1980.
22. Bernard Rosner, "Percentage Points for the RST Many Outlier Procedure", Technometrics **19**, 307-12 (August 1977).
23. Ram B. Jain, "Percentage Points of Many-Outlier Detection Procedures", Technometrics **23**, 71-5 (February 1981).
24. Edward Paulson, "An Optimum Solution to the k-Sample Slippage Problem for the Normal Distribution", Annals of Mathematical Statistics **23**, 610-6 (December 1952).
25. C. P. Quesenberry and H. A. David, "Some Tests of Outliers", Biometrika **48**, 279-90 (December 1961).
26. E. S. Pearson, "Tables of Percentage Points of $\sqrt{b_1}$ and b_2 in Normal Samples; A Rounding Off", Biometrika **52**, 282-5 (June 1965).
27. M. P. Mulholland, "On the Null Distribution of $\sqrt{b_1}$ for Samples of Size at Most 25, With Tables", Biometrika **64**, 401-9 (August 1977).
28. H. A. David, "Revised Upper Percentage Points of the Extreme Studentized Deviate from the Sample Mean", Biometrika **43**, 449-51 (December 1956).
29. M. Halperin, S. W. Greenhouse, J. Cornfield, and Julia Zalokar, "Tables of Percentage Points for the Studentized Maximum Absolute Deviation in Normal Samples", Journal of the American Statistical Association **50**, No. 269, 185-95 (March 1955).
30. K. R. Nair, "Distribution of the Extreme Deviate from the Sample Mean", Biometrika **35**, 118-44 (June 1948).
31. Douglas M. Hawkins and André F. Perold, "On the Joint Distribution of Left- and Right-Sided Outlier Statistics", Utilitas Mathematica **12**, Winnipeg, Canada 129-43 (November 1977).
32. Peter J. Huber, "Robust Statistics: A Review, Special Invited Paper: The 1972 Wald Lecture", Annals of Mathematical Statistics **43**, 1041-67 (August 1972).
33. Robert V. Hogg, "An Introduction to Robust Estimation", *Robustness in Statistics*, Robert L. Lavner and Graham N. Wilkinson, Eds., Academic Press, New York, NY, 1979.
34. S. S. Shapiro and M. B. Wilk, "An Analysis of Variance Test for Non-Normality (Complete Samples)", Biometrika **52**, 591-611 (December 1965).
35. Edwin H. Chen, "The Power of the Shapiro-Wilk W Test for Normality in Samples from Contaminated Normal Distributions", Journal of American Statistical Association **66**, 760-2 (December 1971).
36. S. S. Shapiro and R. S. Francia, "An Approximate Analysis of Variance Test for Normality", Journal of American Statistical Association **67**, 215-6 (March 1972).

37. Cuthbert Daniel, "Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments", Technometrics **1**, 311-42 (May 1959).
38. Herman Chernoff and Gerald J. Lieberman, "Use of Normal Probability Paper", Journal of American Statistical Association **49**, 778-85 (December 1954).
39. Herman Chernoff and Gerald J. Lieberman, "The Use of Generalized Probability Paper for Continuous Distributions", Annals of Mathematical Statistics **27**, 806-16 (September 1956).
40. D. N. Sparks, "Half-Normal Plotting", Applied Statistics **19**, 192-6 (1970).
41. Douglas A. Zahn, "Modifications of and Revised Critical Values for the Half-Normal Plot", Technometrics **17**, 189-200 (May 1975).
42. M. B. Wilk, R. Gnanadesikan, and M. J. Huyett, "Probability Plots for the Gamma Distribution", Technometrics **4**, 1-20 (February 1962).
43. M. B. Wilk and R. Gnanadesikan, "Probability Plotting Methods for the Analysis of Data", Biometrika **55**, No. 1, 1-17 (March 1968).
44. P. Prescott, "Examination of the Behavior of Tests for Outliers When More Than One Outlier is Present", Applied Statistics **27**, 10-25 (1978).
45. P. Prescott and R. V. Hogg, "Trimmed and Outer Means and Their Variances", The American Statistician **31**, 156-7 (November 1977).
46. Frank J. Anscombe, "Rejection of Outliers", Technometrics **2**, 123-47 (May 1960).
47. Janet Dixon Elashoff, "A Model for Quadratic Outliers in Linear Regression", Journal of American Statistical Association **67**, 478-85 (June 1972).
48. Jonas H. Ellenberg, "The Joint Distributions of the Standardized Least Squares Residuals from a General Linear Regression", Journal of American Statistical Association **68**, 941-3 (December 1973).
49. Richard F. Green, "Outlier-Prone and Outlier-Resistant Distributions", Journal of the American Statistical Association **71**, 502-05 (June 1976).
50. V. Barnett and T. Lewis, *Outliers in Statistical Data*, John Wiley & Sons, New York, NY, 1979.

BIBLIOGRAPHY

Frank J. Anscombe, "Examination of Residuals", *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley **1**, 1-36 (1961).

Barry C. Arnold, "Schwarz, Regression, and Extreme Deviance", The American Statistician **28**, 22-3 (February 1974).

R. P. Bland, R. D. Gilbert, C. H. Kapadia, and D. B. Owen, "On the Distributions of the Range and Mean Range for Samples from a Normal Distribution", Biometrika **53**, 245-8 (June 1966).

A. P. Basu, "On Some Hypotheses Relating to the Exponential Distribution When Some Outliers are Present", Journal of the American Statistical Association **60**, 548-59 (June 1965).

Victor Chew, *Tests for the Rejection of Outlying Observations*, RCA Systems Analysis Technical Memorandum No. 64-7, Patrick Air Force Base, FL, December 1964.

Meyer Dwass, "The Extreme Deviations Inequality", The American Statistician **29**, 108 (Letter to the Editor) (May 1975).

Douglas M. Hawkins, "The Detection of Errors in Multivariate Data Using Principal Components", Journal of American Statistical Association **69**, 340-4 (June 1974).

J. A. John and N. R. Draper, "On Testing for Two Outliers or One Outlier in Two-Way Tables", Technometrics **20**, 69-78 (February 1978).

N. L. Johnson, "Approximations to the Probability Integral of the Distribution of Range", Biometrika **39**, 417-9 (December 1952).

Genshiro Kitagawa, "On the Use of AIC for the Detection of Outliers", Technometrics **21**, 193-9 (May 1979).

W. H. Kruskal, "Some Remarks on Wild Observations", Technometrics **2**, 1-3 (February 1960).

R. G. McMillan, "Tests for One or Two Outliers in Normal Samples With Unknown Variance," Technometrics **13**, 87-100 (February 1971).

Frank Proschan, "Testing Suspected Observations", Industrial Quality Control **XIII**, 14-9 (January 1957).

CHAPTER 4

SELECTED TOPICS IN ESTIMATION, THE COMMON STATISTICAL TESTS OF SIGNIFICANCE, AND THE CHOICE OF PERCENTAGE POINTS

Some specially selected topics for the practicing statistician are discussed in this chapter. These include

1. *Unbiased estimation of the normal population standard deviation*
2. *The sample range*
3. *The sample mean deviation*
4. *The concept of mean square error*
5. *Some moment properties of distributions*
6. *The chi-square distribution and its relation to the binomial and Poisson distributions*
7. *Confidence bounds on the unknown normal population standard deviation (sigma)*
8. *The approximate chi-square distribution*
9. *The Snedecor-Fisher variance ratio distribution*
10. *Tests for homogeneity of population variances or homoscedasticity*
11. *Student's t distribution for a single sample and for two samples*
12. *Special approximations to Student's t*
13. *The Behrens–Fisher problem*
14. *Special use of an experimental design to rate or rank proposals*
15. *Combination of probabilities from independent experiments*
16. *Choice of significance levels for multiple tests*
17. *Brief introduction to the field of multiple comparison procedures.*

Statistical tables of percentage points that the analyst will often use are included, and a variety of examples illustrating the theory presented is recorded.

4-0 LIST OF SYMBOLS

A_i = designation for the ith event

A_{ij} = score or rating by the ith rater on the jth proposal

$A_{i.}$ = summation with respect to j

$A_{xx} = n\Sigma x^2 - (\Sigma x)^2$

$A_{.j}$ = summation with respect to i

$A_{..}$ = summation of ratings over both i and j

$\bar{A}_{i.}$ = mean of ratings given by the ith rater on all proposals

$\bar{A}_{.j}$ = mean of ratings by all raters on jth proposal

$\bar{A}_{..}$ = mean of ratings by all raters on all proposals

a_1 = constant in Eq. 4-10

a_2 = constant in Eq. 4-11 = $1/c$

C = denominator of Bartlett's F

c = constant in Eq. 4-6

c_n = constant in Eq. 4-5

d_n = mean value of sample range divided by σ

d_s = special form of Student's t for the Behrens-Fisher problem in Eq. 4-124

$E(\ )$ = expected value of ( )

F = Snedecor-Fisher F statistic: a ratio of variances

F_B = Bartlett's F (or Bartlett's statistic)
F_{BK} = Bartlett-Kendall "log ANOVA" F statistic
F_C = Cochran's F or statistic
$F_H = F_{\max}$ = Hartley's maximum F ratio or statistic
$F_\alpha(\nu_1,\nu_2)$ = lower α probability level of F with ν_1 and ν_2 degrees of freedom
$F_{1-\alpha}(\nu_1,\nu_2) = 1/F_\alpha(\nu_2,\nu_1)$ = upper α probability level of F with ν_1 and ν_2 degrees of freedom
FP = F ratio of mean square for proposals to mean square error or residual variance
FR = F ratio of raters to the residual mean square
$f(\ )$ = probability density function (pdf) of ( )
$f(F)$ = probability density function of Snedecor-Fisher F
$f(t)$ = probability density function of variable t
g = tabular value to use Duncan's Multiple Range test
I = confidence interval
I_{ML} = confidence interval of minimum length
I_{SU} = Neyman's shortest unbiased confidence interval
$I_x(p,q)$ = incomplete beta function
K = constant
k = constant due to Cureton in Eq. 4-9
k = number of proposals
k_n = standard error of sample range divided by σ
L = length of confidence interval in Eq. 4-65
L^* = form of Bartlett's statistic
M = numerator of Bartlett's F
MD = mean deviation from mean in Eq. 4-15
ML = maximum likelihood (estimate)
MS = mean square
MSE = mean square error
MSE = mean square for the error or residual variance term ("error of measurement" for the experiment)
$MSE\,(MD)$ = mean square error of the sample mean deviation
$MSE\,(w)$ = mean square error of the sample range
MSP = mean square for the different proposals
MSR = mean square for the raters
m = designates the number of independent tests carried out
m = mean value of Q
m_i = number of sample variances from ith population
max ( ) = maximum of ( )
$N(0,1)$ = indicates a normal distribution with zero mean and unit variance
n = sample size
n = number of raters
n_1 = sample size of "first" sample (drawn from first population)
n_2 = sample size of "second" sample (drawn from second population)
P_j = the jth proposal
$Pr[\ ]$ = probability that or of [ ]
$Pr[x \geq s]$ = chance that x attains s or more successes

p = number of populations sampled

p = chance of success in a single trial

p^* = actual probability attained with use of t^* instead of Student's t

p_i = left area of a probability distribution up to t_i

p_1 = first left tail area

p_2 = second left tail area

$Q = Q(x,y)$ = quadratic form in x and y

q = designates the "Studentized" range, i.e., the range of a sample of observations divided by the standard deviation

$q = 1 - p$ = chance of failure in a single trial

R_i = ith rater

r_i = sample range of ith sample

$\binom{r}{i}$ = combination of r things taken i at a time

$S_1^2 = \sum_{i=1}^{n_1} (x_{i1} - \bar{x}_1)^2$ = sum of squares about the first sample mean

$S_2^2 = \sum_{i=1}^{n_2} (x_{i2} - \bar{x}_2)^2$ = sum of squares about the second sample mean

SS = sum of squares (about proper mean value)

SSE = sum of squares due to residual or error variance

SSP = sum of squares due to proposals

SSR = sum of squares due to raters

SST = total sum of squares

s = number of successes in n trials

s = sample standard deviation

s_x = sample standard deviation in x-direction

s_y = sample standard deviation in y-direction

$s^2 = \Sigma(x_i - \bar{x})^2/(n-1)$ = sample variance based on $(n-1)$ degrees of freedom

$$s^2 = A_{xx}/[n(n-1)] = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)}$$

$$s^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} (x_i - x_j)^2/[2n(n-1)]$$

$(s')^2 = \Sigma(x_i - \bar{x})^2/n$ = sample variance with divisor n

s_i^2 = sample variance for sample of size n_i from ith normal population

s_{ij}^2 = jth sample variance from ith population

$s_{\max}^2$ = maximum sample variance

$s_{\min}^2$ = minimum sample variance

$s_1^2 = S_1^2/(n_1 - 1)$ = sample variance of first sample based on $(n_1 - 1)$ degrees of freedom

$s_2^2 = S_2^2/(n_2 - 1)$ = sample variance of second sample based on $(n_2 - 2)$ degrees of freedom

T = a general sample statistic in Eq. 4-23

t = Wilson-Hilferty transformation or Student's t statistic

t_s = special form of Student's t in Eq. 4-123

t^* = Scott and Smith's modified Student's t in Eq. 4-105

$t_{0.95}^*$ = 95% value or probability level of t^*

t_1 = upper $\alpha/2$ percentage point for $(n_1 - 1)$ degrees of freedom in Eq. 4-125

t_2 = upper $\alpha/2$ percentage point for $(n_2 - 1)$ degrees of freedom in Eq. 4-125

t_α = upper α probability level of Student's t

UA_i = designates the occurrence of at least one of the events A_i

Var ( ) = variance of ( )
v = variance of Q
w = sample range = $x_{nn} - x_{1n}$
x_i = ith observation
x_{in} = ith ordered value or observation in a sample of size n
$\bar{x} = \Sigma x_i/n$ = sample mean
$\bar{x}_1$ = sample mean of first sample
$\bar{x}_2$ = sample mean of second sample
y = general statistical variable
$z = (\ln F)/2$ = Fisher's transformation of F
z = standard or unit normal deviate
$z_{ij} = \ln s_{ij}^2$ = logarithm of jth sample variance from ith population
$z_{i\cdot}$ = ith average of z_{ij}'s in Eq. 4-96
$z_{\cdot\cdot}$ = grand average of z_{ij}'s in Eq. 4-97
$z_{0.95}$ = 1.96 = 95% probability level of normal deviate z
α = probability level < 0.5
$1 - \alpha$ = confidence level or probability
$\alpha_3 = \mu_3/\mu_2^{3/2}$ = coefficient of skewness = $\sqrt{\beta_1}$
$\alpha_4 = \mu_4/\mu_2^2$ = coefficient of kurtosis = β_2
β = amount of bias in an estimate
$\beta(p,q)$ = beta function of p and q
$\sqrt{\beta_1} = \alpha_3$
$\Gamma(\ )$ = gamma function of ( )
δ = divisor to obtain an unbiased estimate
θ = population parameter
$\lambda = np$ = Poisson expectation parameter
μ = true mean
μ_{ij} = true unknown grade or rating for ith rater on jth proposal
μ_r = rth central moment, or rth moment about the mean
$\mu_r' = \mu_r'(\ )$ = rth moment about the origin of ( )
$\mu_{\cdot j}$ = true unknown mean grade or rating for the jth proposal
μ_1 = population mean of first normal population
μ_2 = population mean of second normal population
$\mu_2 = \sigma^2$ = variance
ν = degrees of freedom (df)
$\nu_i = n_i - 1$ = number of degrees of freedom for ith sample
ν_1 = degrees of freedom for first sample
ν_2 = degrees of freedom for second sample
σ_{MD} = standard deviation of the MD
σ_w = standard deviation of the sample range
σ_x = population standard deviation in x-direction
$\sigma_{\bar{x}_1 - \bar{x}_2}$ = standard deviation of the difference in means
σ_y = population standard deviation in y-direction
σ_o = hypothesized value of σ (for the null hypothesis)

σ_1 = population standard deviation of first normal population
σ_2 = population standard deviation of second normal population
σ^2 = population variance
$\sigma^2(\ )$ = variance of ( )
$\hat{\sigma}$ = estimate of σ (usually the optimum estimate)
$\hat{\sigma}^2$ = estimate of the population variance
$\chi^2(2m^2/\nu)$ = approximate chi-square variate with $2m^2/\nu$ degrees of freedom
$\chi^2(\nu)$ = chi-square using ν degrees of freedom
χ^2_a = a lower limit of chi-squared distribution
χ^2_b = an upper limit of chi-squared distribution
χ^2_α = αth probability level or percentage point of chi-square
$\chi^2_{1-\alpha}$ = $(1-\alpha)$th probability level or percentage point of chi-square

4-1 INTRODUCTION

The fundamental problem of statistics is to improve upon or to develop the most powerful and useful methods for the analysis and interpretation of data of all kinds. In Chapter 2 we developed some of the most up-to-date techniques for determining the precision and accuracy of our measuring instruments and for defining these concepts in useful analytical terms. If our measurements are faulty, the correct interpretation or sound inferences from samples become difficult or impossible; this is the reason for studying the precision and accuracy of measurements. In a like manner, it seems logical and basically sound to examine samples (often expensively taken) for outliers, which also may lead to erroneous conclusions or inferences, before we address the problem of refined methods of statistical analyses. It is true that we applied many of the common statistical tests of significance in Chapters 2 and 3 because they were, in fact, necessary to test various hypotheses of importance. Many of the more common statistical tests of significance are found in standard textbooks on statistics. Nevertheless, we must examine more critically many of the problems related to statistical tests of significance, some problems of confidence interval estimation, and the problem of statistical hypothesis testing generally in order to update techniques for the current practicing Army analyst—especially since the five sections of the Engineering Design Handbooks on experimental statistics (Refs. 1-5) appeared in 1962.

Refs. 1-5 contain a wealth of general and specific information concerning statistical techniques—current to 1962—of interest to the practicing Army analyst. These include, for example,

1. Snedecor's F ratio of sample variances to test the equality of normal population variances
2. Student's t statistic for testing the hypothesis concerning whether the population mean for a normal sample has a specified value or the two-sample Student's t for comparing population means
3. Contingency tables and other statistical tests for comparing the true unknown proportions of binomial- or multinominal-type populations
4. Analysis of scientific experiments including factorial experiments
5. Completely randomized blocks and incomplete block designs, Latin squares, Youden squares and other special designs
6. Transformations of data to stabilize variances or to assure normality
7. Some topics in least squares, regression, and curve fitting
8. Confidence intervals
9. Many other useful statistical techniques or procedures for either the new or experienced statistical analyst.

In addition, Section 5 of the *Experimental Statistics* Handbook (Ref. 5) contains many very valuable statistical tables, including some not ordinarily found in standard statistical textbooks. Since this valuable set of statistical methods is already available to the Army statistician, it becomes our main purpose to carry forward some of the more useful and important topics that have been developed during the past 16 yr and to cover some particular topics of current interest as now envisioned for Army applications. Although there will be a minimal amount of repetition with regard to Refs. 1-5, this will be presented and discussed only as

necessary background and as deemed necessary to introduce or carry forward our suggested applications. We will start with some preliminaries concerning the drawing of random samples from a single normal population and will give some results of interest that have either appeared in the statistical literature since about 1962, i.e., the appearance of Refs. 1-5, or at least have not been covered in these handbooks. Ref. 1 gives a good account of elementary concepts.

4-2 PRELIMINARY REMARKS ON SAMPLING A SINGLE NORMAL POPULATION

4-2.1 THE SAMPLE MEAN AND STANDARD DEVIATION

We start with the concept of drawing a single random sample of size n from a normal or Gaussian population with true mean μ and standard deviation σ, or variance σ^2. The observations come in a random and unordered sequence as contrasted to that discussed in Chapter 2, and we designate them as $x_1, x_2, \ldots, x_i, \ldots x_n$. For the present, we will be primarily interested in the sample mean $\bar{x}$, or

$$\bar{x} = \sum_{i=1}^{n} x_i/n \tag{4-1}$$

and the sample variance s^2 based on $(n-1)$ degrees of freedom (df), or

$$s^2 = \sum_{i=1}^{n}(x_i-\bar{x})^2/(n-1) = \frac{n\Sigma x^2 - (\Sigma x)^2}{n(n-1)} = \frac{A_{xx}}{n(n-1)} \tag{4-2}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - x_j)^2/[2n\,(n-1)].$$

We might on some occasions have interest in the sample standard deviation s', which uses the total sample size n instead of the number of df $= (n-1)$, i.e.,

$$s' = [\Sigma(x_i - \bar{x})^2/n]^{1/2} = \sqrt{(n-1)/n}\, s. \tag{4-3}$$

It is well-known, e.g., from standard textbooks on statistics, that $\bar{x}$ is the maximum likelihood (ML) unbiased $[E(\bar{x}) = \mu]$, minimum variance, most efficient estimator of μ the normal population true mean and that the variance of $\bar{x}$ is simply

$$\mathrm{Var}(\bar{x}) = \sigma^2(\bar{x}) = \sigma^2/n. \tag{4-4}$$

The sample variance s^2 based on $(n-1)$ df is the unbiased estimate of the population variance σ^2, or $E(s^2) = \sigma^2$, although the maximum likelihood estimate of σ^2 is s'^2, but it is biased or $E(s'^2) = (n-1)\sigma^2/n$. Concerning estimates of the population standard deviation σ, both s and s' are biased, unfortunately, and involve a ratio of gamma functions. That is to say

$$E(s') = \sqrt{(n-1)/n}\; E(s) = \sqrt{2/n}\;\Gamma\left(\frac{n}{2}\right)\sigma/\Gamma[(n-1)/2] = c_n\sigma \tag{4-5}$$

where

$\Gamma(\ \ )$ = gamma function of $(\ \ )$

c_n = constant depending on the sample size n.

Many writers have used "c_2" instead of "c_n". Here we take

$$E(s) = c\sigma \text{ or } c = \sqrt{n}\, c_n/\sqrt{n-1} \tag{4-6}$$

where

c = constant.

For a discussion of many of the more elementary statistics and their properties—especially as related to the military sciences and the delivery accuracy of weapons, etc.—the reader may consult Ref. 6 and the appendix of it.

The fact that both s and s' are biased estimates of the normal population standard deviation, or parameter σ, has stimulated much thought and study of this problem, especially toward providing simple, accurate approximations of the involved quantity c_n, or ratio of gamma functions in Eq. 4-5. Is there really a simple and accurate approximation of c_n that the statistician can easily remember? This is the kind of problem that may be dormant for many, many years and then sudden interest may cause much investigation. Precisely this happened as late as 1968 when Cureton (Ref. 7) published, as a teaching aid, a table of values for obtaining the unbiased estimate of the normal population sigma in "The Teacher's Corner" of the *American Statistician*. Cureton (Ref. 7) apparently was interested in taking the sum of squares about the sample mean, dividing it by a quantity he calls "k", and then taking the square root to obtain an unbiased estimate of the normal population σ. That is to say,

$$\text{Unbiased Est } \sigma = \hat{\sigma} = \sqrt{\Sigma(x_i - \bar{x})^2/k} \tag{4-7}$$

so that the expected value E is truly unbiased, or

$$E(\hat{\sigma}) = \sigma \tag{4-8}$$

for a normal population. We note also that the relations between our c_n, or our c, and Cureton's k are

$$c_n = \sqrt{k}\ /\sqrt{n} \text{ or } c = \sqrt{k}\ /\sqrt{n-1}\,. \tag{4-9}$$

Cureton gives a very compact table of values for k, presented here as Table 4-1, which, with only 26 line entries, covers all sample sizes up to and beyond $n = 252$. Note that for $n \geq 21$ and for three decimal places the values of k change very slowly and approach the value $(n - 1.5)$. This suggests that $(n - 1.5)$ would be a better divisor than the $(n - 1)$ df, insofar as unbiasedness is concerned; although for $n = 2$ we have that $n - 1.5 = 0.5$ instead of the correct value 0.6366. More will be said of this in the sequel.

TABLE 4-1

VALUES OF k IN $\sqrt{\Sigma(x_i - \bar{x})^2/k}$ TO OBTAIN UNBIASED ESTIMATES OF NORMAL POPULATION σ (Ref. 7)

n	k	n	k
2	0.6366	15	13.509
3	1.571	16	14.509
4	2.546	17	15.508
5	3.534	18	16.508
6	4.527	19	17.507
7	5.522	20	18.507
8	6.519	21-24	$n - 1.494$
9	7.517	25-29	$n - 1.495$
10	8.515	30-37	$n - 1.496$
11	9.513	38-51	$n - 1.497$
12	10.512	52-83	$n - 1.498$
13	11.511	84-251	$n - 1.499$
14	12.510	252 up	$n - 1.500$

Reprinted with permission. Copyright © by the American Statistical Association.

Following Cureton's letter (Ref. 7) by four months, Bolch (Ref. 8) gives a five-decimal place table of values of multipliers, a_1 and a_2, by which to multiply s' and s, respectively, to obtain the unbiased estimate for various sample sizes n of the normal population sigma. Bolch's table (Ref. 8) is included as Table 4-2, and we see that

$$E(a_1 s') = E(s'/c_n) = \sigma^* \tag{4-10}$$

and

$$E(a_2 s) = \sigma. \tag{4-11}$$

Thus Bolch's table, covering many sample sizes, will be useful to analysts or statisticians to obtain unbiased estimates of σ in their work.

Although the rash of letters to the editor of *The American Statistician* on unbiased estimation of the normal population standard deviation continued for some years, Gurland and Tripathi (Ref. 9) showed that a good approximation of Cureton's k, instead of the $(n - 1.5)$, is simply

$$k = n/a_1^2 \approx n - 1.5 + 1/[8\,(n-1)] \tag{4-12}$$

and that the quantity $1/c$ is nearly

$$1/c = a_2 \approx (4n - 3)/\,[4\,(n-1)]. \tag{4-13}$$

And from Eq. 4-11, $E(a_2 s) = \sigma$. Even for $n = 2$, the exact value of Cureton's k is 0.6366, whereas Eq. 4-12 gives 0.6250, and for larger n Eq. 4-12 approaches the exact value of k very rapidly.

When $n = 2$, the exact value of $1/c$ is 1.2533, whereas Eq. 4-13 gives $1/c = 1.2500$, and again the differences disappear rapidly with larger sample sizes n.

With reference to Eq. 4-12, Bhoj (Ref. 10) indicates that an improvement in the accuracy of k may be obtained by using

$$k \approx n - 1.5 + 1/[8\,(n - 1.45)]. \tag{4-14}$$

The sample variance, s^2 of Eq. 4-2, is an unbiased estimate of the population variance σ^2 of any continuous distribution having finite σ^2, whereas the bias in s and s' depends on the distribution of individuals in the population sampled.

With this updating of accomplishments on the sample standard deviation for the normal population, for our purposes we will record only two other measures of dispersion for univariate samples—the sample mean deviation (MD) and the sample range. A good coverage of both univariate and bivariate measures of dispersion for the Army analyst, including quantification of their relative efficiencies and other properties, may be found in Ref. 6.

4-2.2 THE SAMPLE MEAN DEVIATION

The mean deviation MD of the sample, or the mean absolute deviation as it is often called, is defined by

$$MD = \sum_{i=1}^{n} |x_i - \bar{x}|/n. \tag{4-15}$$

Thus the MD is simply the average of the unsigned (all positive) deviations of the observations from the sample mean.

---

*$E(s') \approx \sigma\,[1 - 3/(4n) - 7/(32n^2) - 9/(128n^3)]$, $a_1 \approx \dfrac{(n - 0.25)}{(n-1)}$

$\sigma_s^2 \approx [\sigma^2/(2n)]\,[1 - 1/(4n) - 3/(8n^2)]$

TABLE 4-2

VALUES OF a_1 AND a_2 SUCH THAT $a_1 s'$ AND $a_2 s$ ARE UNBIASED ESTIMATES OF THE NORMAL POPULATION STANDARD DEVIATION (Ref. 8)

n	a_1	a_2	n	a_1	a_2
2	1.77245	1.25331	34	1.02275	1.00760
3	1.38198	1.12838	35	1.02209	1.00738
4	1.25331	1.08540	36	1.02145	1.00717
5	1.18942	1.06385	37	1.02086	1.00697
6	1.15124	1.05094	38	1.02029	1.00678
7	1.12587	1.04235	39	1.01976	1.00660
8	1.10778	1.03624	40	1.01925	1.00643
9	1.09424	1.03166	41	1.01877	1.00627
10	1.08372	1.02811	42	1.01831	1.00612
11	1.07532	1.02527	43	1.01788	1.00597
12	1.06844	1.02296	44	1.01746	1.00583
13	1.06272	1.02103	45	1.01706	1.00570
14	1.05788	1.01940	46	1.01668	1.00557
15	1.05373	1.01800	47	1.01632	1.00545
16	1.05014	1.01679	48	1.01597	1.00533
17	1.04700	1.01574	49	1.01564	1.00522
18	1.04423	1.01481	50	1.01532	1.00511
19	1.04176	1.01398	60	1.01272	1.00425
20	1.03956	1.01324	70	1.01088	1.00363
21	1.03758	1.01257	80	1.00950	1.00317
22	1.03579	1.01197	90	1.00843	1.00281
23	1.03416	1.01142	100	1.00758	1.00253
24	1.03267	1.01093	110	1.00688	1.00230
25	1.03130	1.01047	120	1.00630	1.00210
26	1.03005	1.01005	130	1.00582	1.00194
27	1.02888	1.00965	140	1.00540	1.00180
28	1.02783	1.00931	150	1.00503	1.00168
29	1.02682	1.00897	160	1.00472	1.00158
30	1.02590	1.00866	170	1.00445	1.00149
31	1.02503	1.00836	180	1.00420	1.00141
32	1.02423	1.00810	190	1.00395	1.00130
33	1.02347	1.00784	200	1.00378	1.00127

Reprinted with permission. Copyright © by the American Statistical Association.

For a normal population, the expected, or mean, value of the MD is

$$E(MD) = \sqrt{2(n-1)/(n\pi)}\,\sigma \to 0.7979\sigma \tag{4-16}$$

and, as indicated for large sample size n, approaches $\sqrt{2/\pi}\,\sigma = 0.7979\sigma$. Thus the MD is also a biased estimate of σ for a normal distribution, and approaches a value about 0.2σ less than the normal population sigma.

It has been shown by Fisher (Ref. 11) that the standard error σ_{MD} of the MD in samples of size n from a normal universe is

$$\sigma_{MD} = \langle\, 2\,(n-1)\,\{\,(\pi/2) + \sqrt{n(n-2)} - n + \mathrm{Sin}^{-1}\,[1/(n-1)]\}/(n^2\pi)\,\rangle^{1/2}\sigma. \tag{4-17}$$

In Table 4-3 we give the mean values and standard deviations of the MD for samples of size $n = 2\,(1)\,20$ and also the 95% probability level or percentage points of the MD. More details on the MD may be found in Ref. 6.

TABLE 4-3

MEAN VALUES AND STANDARD DEVIATIONS OF THE SAMPLE MEAN DEVIATION (Ref. 6)

Sample Size n	Mean Value of MD $E(MD)/\sigma$	Reciprocal of Mean Value Coefficient	Standard Deviation of MD σ_{MD}/σ	95% Probability Limit
2	0.5642	1.772	0.4263	1.39
3	0.6515	1.535	0.3419	1.28
4	0.6910	1.447	0.2970	1.22
5	0.7137	1.401	0.2663	1.19
6	0.7284	1.373	0.2436	1.16
7	0.7387	1.354	0.2258	1.14
8	0.7464	1.340	0.2115	1.12
9	0.7523	1.329	0.1996	1.10
10	0.7569	1.321	0.1894	1.09
11	0.7608	1.314	0.1807	1.07
12	0.7639	1.309	0.1731	1.06
13	0.7666	1.304	0.1664	1.05
14	0.7689	1.301	0.1604	1.04
15	0.7708	1.297	0.1550	1.04
16	0.7725	1.294	0.1501	1.03
17	0.7741	1.292	0.1457	1.02
18	0.7754	1.290	0.1416	1.02
19	0.7766	1.288	0.1378	1.01
20	0.7777	1.286	0.1344	1.01

4-2.3 THE SAMPLE RANGE

We made use of the sample range in discussing bounds and as a test of the lowest and highest sample observations in pars. 3-2.1, 3-2.2, and 3-5.3. If we now designate the ordered sample observations as

$$x_{1n} \leq x_{2n} \leq \cdots \leq x_{in} \leq \cdots \leq x_{nn} \tag{4-18}$$

where

$x_{in} = i$th ordered value or observation is a sample size n

the sample range w becomes

$$w = x_{nn} - x_{1n}. \tag{4-19}$$

Clearly, the sample range depends markedly on the sample size n, and w increases with increasing n.

It has been customary by many writers to designate the expected or mean value of the sample range by

$$E(w) = d_n\sigma \tag{4-20}$$

showing that the factor or coefficient d_n, the multiplier of the normal population sigma, depends on the sample size n. Moreover, it has been statistical practice to designate the variance of w as

$$\text{Var}(w) = \sigma^2(w) = E(w - d_n\sigma)^2 = k_n^2\sigma^2 \tag{4-21}$$

whereas the standard error of w is

$$\sigma(w) = k_n\sigma \tag{4-22}$$

where

k_n = standard error of sample range divided by σ.

In Table 4-4 we give the quantities d_n, $1/d_n$, and k_n for the range constants and samples of size 2(1)20 drawn from a normal population. One may note that d_n increases rather rapidly with increasing sample size and that k_n decreases slowly, which indicates a moderate improvement in precision with increasing n. In Table 4-4 we also give the 95% probability values for the sample range in case this might be of some use to the analyst.

Let us now turn to Example 4-1 concerning the sample standard deviation, the sample mean deviation, and the sample range.

Example 4-1:

Given the 11 muzzle velocities: 1480, 1501, 1510, 1499, 1492, 1509, 1500, 1502, 1498, 1479, and 1490 in ft/s for rounds fired from a 155-mm Howitzer, calculate the expected muzzle velocity of the weapon and the unbiased estimate of the population sigma using (1) the sample standard deviation, (2) the sample MD, and (3) the sample range.

The sample standard deviation based on $(n - 1) = 10$ df is from Eq. 4-2

$$s = 10.25.$$

By using Eq. 4-11 and Table 4-2 from $n = 11$, the unbiased estimate of σ is

$$\text{est}\sigma = (1.02527)\ (10.25) = 10.51 \text{ ft/s}.$$

TABLE 4-4

MEAN VALUES AND STANDARD DEVIATIONS OF THE SAMPLE RANGE w (Ref. 6)

Sample Size n	Mean Value $d_n = E(w)/\sigma$	Reciprocal of Mean Value Coefficient	Standard Deviation $k_n = \sigma_w/\sigma$	95% Probability Limit
2	1.128	0.8862	0.8525	2.77
3	1.693	0.5908	0.8884	3.31
4	2.059	0.4857	0.8798	3.63
5	2.326	0.4299	0.8641	3.86
6	2.534	0.3946	0.8480	4.03
7	2.704	0.3698	0.8332	4.17
8	2.847	0.3512	0.8198	4.29
9	2.970	0.3367	0.8078	4.39
10	3.078	0.3249	0.7971	4.49
11	3.173	0.3152	0.7873	4.55
12	3.258	0.3069	0.7785	4.62
13	3.336	0.2998	0.7704	4.69
14	3.407	0.2935	0.7630	4.74
15	3.472	0.2880	0.7562	4.80
16	3.532	0.2831	0.7499	4.85
17	3.588	0.2787	0.7441	4.89
18	3.640	0.2747	0.7386	4.93
19	3.689	0.2711	0.7335	4.97
20	3.735	0.2677	0.7287	5.01

The expected muzzle velocity is $\bar{x} = 1496.36$ ft/s.
From Eq. 4-15 the sample MD is

$$MD = \Sigma|x_i - \bar{x}|/11 = 8.083.$$

By using the reciprocal of the mean value coefficient for $n = 11$ from Table 4-3, we get

$$\text{est}\sigma = (1.314)\ (8.083) = 10.62 \text{ ft/s}$$

a slightly larger value.

Finally, the sample range is

$$w = 1510 - 1479 = 31 \text{ ft/s}$$

and by multiplying this by the value of 0.3152 for $n = 11$ in Table 4-4 or by dividing 31 by 3.173, we obtain (Eq. 4-20)

$$\text{est}\sigma = 9.77 \text{ ft/s}$$

which turns out to be the smallest of the estimates of σ.

The sample range, of course, is the easiest and quickest sample statistic from which to calculate and to estimate the normal population sigma, whereas the sample standard deviation results in a more complex type of calculation. It can be said, however, that with modern pocket electronic calculators the striking difference in effort nearly disappears—especially if we also consider the matter of efficiency of estimators. We discuss this next along with the evercontinuing controversy concerning the use of biased or unbiased estimators in practice.

4-2.4 BIASED OR UNBIASED ESTIMATORS AND EFFICIENCY

The differences in unbiased estimators due to sample size, also differences in ease of computation of the sample range, and even the sample mean deviation having been noted, it certainly becomes of interest to discuss biased versus unbiased estimates further. Also we would like to get some idea as to the relative efficiency of different estimators of the same population parameter, in this case the standard deviation.

Generally, if we are interested in estimating a population parameter, for example, θ—which may be a mean, standard deviation, variance, or other parameter—and we use a sample statistic T, then T will be an *unbiased* estimator of θ if

$$E(T) = \theta. \tag{4-23}$$

On the other hand, if

$$E(T) = \theta + \beta = \delta\theta \tag{4-24}$$

where

β = amount of bias in an estimate, $\beta \neq 0$
δ = divisor to obtain an unbiased estimate, $\delta \neq 1$
$= 1 + \beta/\theta$, $\theta \neq 0$

then it is said that T is a *biased* estimate of the parameter θ. Should we really worry about biased estimators, especially in practice? The answer would seem to be yes, and we cite an example. Examining Table 4-4, we see that the sample statistic w, or the range, is a very biased estimate of the normal population σ. For a sample of

size $n = 2$ it is about 13% higher than the true σ on the average, and for a sample of size $n = 20$ the sample range averages to be about 3.74 times σ! This would seem to be rather intolerable.

If we examine the sample standard deviations s and s', then—by means of Eqs. 4-10 and 4-11 and Table 4-2—we see that they converge rather rapidly to the true population sigma with increasing sample size, i.e., both becoming unbiased. As shown by Eq. 4-16, the sample MD, however, never gets larger than about 0.8σ! Why not then account for and correct for such differences in practice since the bias usually depends on sample size?

If we have several sample statistics that may be used to estimate the same population parameter, some criterion has to be decided upon to select the "best" estimator. We could use the sample statistic that has the least bias, for example, or we could recommend that one having the smallest variance, or the one having the smaller "mean square error" (MSE) (see Eq. 4-26), etc. If we refer to the MD for $n = 10$, we see from Table 4-3 that it has a standard error of 0.1894σ, whereas for the same sample size we have from Table 4-4 that the range has a standard error of 0.7971σ, so that the sample range seems "four times as bad as the sample MD"! However, is this really an accurate analysis since we have not corrected for biases? This type of problem leads us to the concept of MSE. The MSE of a biased estimate T of θ, a population parameter whose expected value is

$$E(T) = \theta + \beta \tag{4-25}$$

where β is the bias, is

$$MSE = E(T - \theta)^2 = \mathrm{Var}(T) + \beta^2. \tag{4-26}$$

The MSE of the sample MD or $MSE\,(MD)$ is

$$MSE\,(MD) = \mathrm{Var}(MD) + [E(MD) - \sigma]^2 \tag{4-27}$$

where

$E(MD)$ = mean value of the sample mean deviation

and the MSE of the sample range $MSE(w)$ is

$$MSE(w) = [(k_n)^2 + (d_n - 1)^2]\,\sigma^2. \tag{4-28}$$

To amplify further the concept of MSE, consider a normal population and the problem of determining the best constant K in

$$\Sigma(x_i - \bar{x})^2/K \tag{4-29}$$

to obtain a very efficient estimate of the population variance σ^2. We already know that if $K = n - 1$, Eq. 4-29 becomes the unbiased estimate of σ^2. However, if we were to choose K so that the MSE (Eq. 4-26) is a minimum, it can be shown that

$$K = n + 1 \tag{4-30}$$

which certainly makes Eq. 4-29 a biased estimate of σ^2.

We will apply the MSE concept in Example 4-2 to the sample range and the sample mean deviation and will show numerically that its worth is questionable for large biases.

Example 4-2:

For a sample of size $n = 11$, determine the MSE of the MD and also the MSE of the sample range. Discuss whether this numerical comparison provides a satisfactory way to select the superior estimator of sigma.

From Eq. 4-27 and Table 4-3, we see the MSE of the MD is

$$MSE(MD) = [(0.1807)^2 + (0.2392)^2]\sigma^2 = 0.0899\sigma^2$$

where σ^2 is the normal population variance.

From Eq. 4-28 and Table 4-4, however, we have for the range w that

$$MSE(w) = [(0.7873)^2 + (3.173 - 1.000)^2]/\sigma^2 = 5.342\sigma^2!$$

Hence for the sample range we obtain an unusually large MSE relatively speaking, but this is due primarily to the large bias in the expected value of the sample range. Admittedly, we have chosen a rather severe example concerning the usefulness of the MSE criterion, but it does show that the MSE may leave much to be desired. This brings us to a much more reasonable and perfectly satisfactory technique for comparing sample statistic efficiencies on practical grounds.

A way out of this dilemma is to make the competitive estimates unbiased so they will have the same mean value and then to compare the variances, or precisions, of the different estimators and select the one with the smallest variance. In other words, for any general estimator T, which is a a biased estimate of θ as indicated by $\delta \neq 1$ in Eq. 4-24, then obviously

$$E(T/\delta) = \theta \tag{4-31}$$

precisely, and the variance of T/δ is therefore

$$\text{Var}(T/\delta) = (1/\delta^2)\text{Var}(T) = \sigma^2(T)/\delta^2. \tag{4-32}$$

Thus we see that the standard error of T/δ, the unbiased estimator, is simply the usual standard deviation of T, the biased estimator, divided by its mean value.

Returning to Example 4-2, we may now compare the relative precisions, or "efficiencies", of the MD and the sample range. Thus for $n = 11$ the relative precision of the MD is simply

$$\sigma(MD/\text{mean value}) = 0.1807/0.7608 = 0.2375$$

and that for the sample range is

$$\sigma(w/d_n) = k_n/d_n = 0.7873/3.173 = 0.2481.$$

In other words, there is practically no difference whatever in the relative efficiencies of the MD and range for $n = 11$ and, hence, little choice unless the range is inflated by outliers (Chapter 3).

The interested reader will find a large number of comparisons of relative precision of unbiased estimates of both univariate and bivariate dispersion population parameters in Table 9 of Ref. 6. For example, it is shown there that, when using the range, a sample of size of $n = 17$ is required to obtain the same precision for estimating the normal population sigma as for a sample of size $n = 13$ when the sample standard deviation is used.* It is only for samples of size two that the standard deviation, the range, and the MD all have equal precision.

In summary, there is no reason why we cannot always deal with unbiased estimators by simply correcting for bias and then use the estimator that is the more precise one. In fact, for nearly all of the sample statistics, the amount of bias will depend on the sample size itself and thereby will bring in an additional complication unless an adjustment is made. Finally, on practical grounds we will usually desire to correct for any sample bias since we are almost always dealing with small size samples.

4-3 SOME MOMENT PROPERTIES

In dealing with the distributional properties of sample statistics, it is quite natural to obtain moments about the origin. However, once the mean of the distribution is determined or estimated, it is the central moments in

*Note that s based on $(n - 1)$ df and s'—both of which use the sample size n—are equivalent in relative precision.

which we are primarily interested. In fact, the second, third, and fourth central moments lead to the variance, the skewness (nonsymmetrical), and kurtosis (peakedness), respectively—properties of the distribution. It is for this reason that we must record the relations between certain of the central moments and the corresponding moments about the origin.

If we define the rth moment about the mean of any general statistical variable y as μ_r, we have the computational equation

$$\begin{aligned}\mu_r &= E[y - E(y)]^r \\ &= \sum_{i=0}^{r} (-1)^i \binom{r}{i} (\mu_1')^i (\mu_{r-i}') \end{aligned} \tag{4-33}$$

where

μ_r' = rth moment about the origin
$\binom{r}{i}$ = combination of r things taken i at a time

which gives central moments in terms of moments about the origin.

The second, third, and fourth central moments in terms of moments about the origin from Eq. 4-33 are, respectively,

$$\mu_2 = \mu_2(y) = \text{Variance} = \mu_2' - (\mu_1')^2 \tag{4-34}$$

$$\mu_3 = \mu_3' - 3(\mu_2')(\mu_1') + 2(\mu_1')^3 \tag{4-35}$$

$$\mu_4 = \mu_4' - 4(\mu_3')(\mu_1') + 6(\mu_2')(\mu_1')^2 - 3(\mu_1')^4. \tag{4-36}$$

Finally, and as a matter of record, Eq. 4-33 may be inverted to give moments about the origin in terms of moments about the mean; the general equation is

$$\begin{aligned}\mu_r' &= E\{[y - E(y)] + E(y)\}^r \\ &= \sum_{i=0}^{r} \binom{r}{i} (\mu_{r-i})(\mu_1')^i. \end{aligned} \tag{4-37}$$

The coefficient of skewness α_3 of any distribution is defined as

$$\alpha_3 = \mu_3/\mu_2^{3/2} = \mu_3/\sigma^3 \tag{4-38}$$

and the kurtosis, or degree of peakedness, coefficient α_4 by

$$\alpha_4 = \mu_4/\mu_2^2 = \mu_4/\sigma^4 \tag{4-39}$$

4-4 THE CHI-SQUARE DISTRIBUTION AND SOME OF ITS USES

4-4.1 THE CHI-SQUARE DISTRIBUTION

Although the normal or Gaussian distribution has long taken the central role in much of the entire field of statistics, the chi-square distribution is perhaps next in importance. In fact, the chi-square distribution may be derived from many standpoints and for both continuous or discrete random variables. Here we will make use of chi-square primarily in terms of the observational sum of squares about the sample mean, especially since we will be interested in testing hypotheses about the size of the normal population variance and in placing confidence intervals on it. Chapter 4, Ref. 1, discusses the problem of comparing the variability of performance of different processes, products or sources, and gives some examples on uses of the theory covered therein. Here we will consider first the sampling of a single normal population and proceed in the direction of updating or expanding the coverage of Ref. 1.

It is well-known for a single random sample taken from a normal population that the SS about the sample mean follows the chi-square distribution with df equal to one less than the sample size. That is,

$$\Sigma(x_i - \bar{x})^2/\sigma^2 = \chi^2\,(n-1). \tag{4-40}$$

The probability density function (pdf) for the random variable χ^2 with ν df is given by

$$f(\chi^2) = \{1/[2^{\nu/2}\Gamma(\nu/2)]\}\,(\chi^2)^{(\nu-2)/2}\,e^{-(\chi^2/2)} \tag{4-41}$$

and χ^2 has a lower limit of zero and an upper limit of plus infinity. The pdf of χ^2 is skewed to the right, especially for small numbers of df ν. When ν becomes large ($\nu \geq$ about 30), the curve becomes more bell shaped and finally approaches the normal, or Gaussian, form.

The mean, variance, and all of the moments of $\chi^2\,(\nu)$ are easily found. In fact, the rth moment μ_r' about the origin of χ^2 is simply

$$\mu_r' = E[(\chi^2)^r] = 2^r\Gamma[r + (\nu/2)]/\Gamma(\nu/2) \tag{4-42}$$

from which all of the central moments, or moments about the mean, are determined. The mean of χ^2 is the number of df or

$$\mu_1' = E(\chi^2) = \nu \tag{4-43}$$

and the variance of χ^2 is simply twice the number of df, i.e.,

$$\mathrm{Var}(\chi^2) = E(\chi^2 - \nu)^2 = 2\nu. \tag{4-44}$$

For the chi-square distribution the coefficient of skewness is

$$\alpha_3 = \alpha_3\,(\chi^2) = 2^{3/2}/\sqrt{\nu}. \tag{4-45}$$

Eq. 4-45 shows that, as the number of df ν increases, $\alpha_3 \to 0$ and the skewness disappears, thus bringing about symmetry of the ultimate or large sample chi-square distribution.

For the chi-square distribution the kurtosis coefficient α_4 is given by

$$\alpha_4 = \alpha_4\,(\chi^2) = 3 + 12/\nu \tag{4-46}$$

showing that for large numbers of df ν, $\alpha_4 \to 3$, which is the value for the normal distribution.

Our primary interest at this point is to discuss and to illustrate some of the special uses of the chi-square distribution based on sampling a single normal distribution, especially the identical quantities

$$(n-1)s^2/\sigma^2 = ns'^2/\sigma^2 = \Sigma(x_i - \bar{x})^2/\sigma^2 = \chi^2(n-1) \tag{4-47}$$

all of which follow the chi-square distribution with $\nu = (n-1)$ df.

Since we know the moments of χ^2 from Eq. 4-42, we may obtain the important moments of s^2 and $(s')^2$. For example, referring to Eqs. 4-40 and 4-43, we see that

$$E(s^2) = (n-1)\sigma^2/(n-1) = \sigma^2 \tag{4-48}$$

or s^2 is unbiased, and from Eq. 4-44

$$\mathrm{Var}(s^2) = 2\,(n-1)\,[\sigma^2/(n-1)]^2 = 2\sigma^4/(n-1) \tag{4-49}$$

which was used in Chapter 2.

Since the chi-square distribution is of the form given by Eq. 4-41, then it is seen that the pdf of chi (χ) is easily obtained by a transformation of variables or by correcting the differential element. This leads to any moment of s or s'. In fact, the rth moment of s', for example, about the origin is

$$E(s')^r = (2\sigma^2/n)^{r/2}\Gamma\left(\frac{n+r-1}{2}\right)/\Gamma\frac{(n-1)}{2}. \tag{4-50}$$

For the mean of s' we put $r = 1$ and obtain

$$E(s') = \sqrt{2}\sigma\Gamma(n/2)/\{\sqrt{n}\,\Gamma[(n-1)/2]\} = c_n\sigma \tag{4-51}$$

$$\approx [1 - 3/(4n) - 7/(32n^2) - 9/(128n^3)]\sigma.$$

The variance of s' is easily found to be

$$\mathrm{Var}(s') = [(n-1)/n - c_n^2]\sigma^2. \tag{4-52}$$

4-4.2 CHI-SQUARE, BINOMIAL, AND POISSON DISTRIBUTION RELATIONSHIPS

It is well-known that for a discrete binomial random variable $x = 0, 1, 2, \ldots, n$ successes (or failures) and also for the chance of success (failure) in a single trial equal to p, the chance of s* or more successes in n preset, fixed trials is given by

$$Pr\,[x \geq s] = \sum_{x=s}^{n} \binom{n}{x} p^x (1-p)^{n-x}. \tag{4-53}$$

This binomial sum is tabulated in many available publications, including the very useful tables in Ref. 12. For reference purposes the useful moment properties of a binomial random variable are

$$\text{Mean} = E(x) = np \tag{4-54}$$
$$\text{Variance} = \sigma^2(x) = npq, \quad (q = 1 - p) \tag{4-55}$$
$$\text{Skewness} = \alpha_3 = (q-p)/\sqrt{npq}). \tag{4-56}$$
$$\text{Kurtosis} = \alpha_4 = 3 + (1 - 6pq)/(npq). \tag{4-57}$$

For small $p \leq$ about 0.10 and np approaching a fixed limit $\lambda = np$, the binomial distribution sum of Eq. 4-53 may be approximated by the Poisson sum

$$Pr\,[x \geq s] \approx \sum_{x=s}^{\infty} e^{-\lambda}\lambda^x/x!. \tag{4-58}$$

Furthermore, the mean and variance of the Poisson distribution are, respectively

$$\text{Mean} = E(x) = \lambda \tag{4-59}$$

which also is equal to the variance, i.e.,

$$\text{Variance} = \sigma^2(x) = \lambda$$

as evidenced by replacing np by λ and $q \approx 1$ in Eqs. 4-54 and 4-55. Thus and in summary, the binomial approaches its Poisson limit when the chance of success (failure) in a single trial is very small.

The very useful relationship between the Poisson and the chi-square distributions for ν df is expressed as

*This s for the number of "successes" in n trials is not to be confused with the sample standard deviation s used previously.

$$\sum_{x=0}^{s-1} \lambda^x \exp(-\lambda)/x! = \int_{\chi^2}^{\infty} \exp(-\chi^2/2)(\chi^2)^{(\nu/2)-1} d\chi^2/[2^{(\nu/2)}\Gamma(\nu/2)] \quad (4\text{-}60)$$

where

$$s = \nu/2 \quad (4\text{-}61)$$

$$\lambda = \chi^2/2. \quad (4\text{-}62)$$

It is due to the relationship in Eq. 4-60 that the probability integral of the chi-square distribution and that of the Poisson are often tabulated together, as in Ref. 13, Table 7, p. 122-9.

4-4.3 SIGNIFICANCE TEST FOR THE SIZE OF A NORMAL POPULATION VARIANCE

Since in the form used here chi-square is expressible as the ratio of the sample sum of squares to the normal population variance, one may test the hypothesis concerning the actual size of the unknown population variance σ^2. This is done by calculating the sum of squares about the sample mean, dividing the result by the hypothesized value of the normal population variance, and then referring this ratio to a table of percentage points of the chi-square distribution. We illustrate this by Example 4-3.

Example 4-3:

Refer to the data on the sample of 11 muzzle velocities (MV) for the 155-mm Howitzer of Example 4-1; make a judgment concerning whether the unknown normal population σ is 15 ft/s.

The sample SS about the sample mean is

$$\Sigma(x_i - \bar{x})^2 = 1050.55$$

and taking $\sigma = 15$, the observed value of χ^2 for 10 df is

$$\chi^2 = 1050.55/(15)^2 = 4.67.$$

To test whether $\sigma = 15$, let us adopt the two-sided test (5% in each tail) or 10% level of significance, and we see that

$$\chi^2_{0.05}(10) = 3.94$$

$$\chi^2_{0.95}(10) = 18.31$$

from, for example, Table A-3 of Ref. 5, which we include here as Table 4-5. Thus the observed value of the sample SS is not quite small enough to reach the $\chi^2_{0.05}$ of 3.94, or large enough to reach the $\chi^2_{0.95} = 18.31$, and hence to conclude that the population sigma is not $\sigma = 15$ ft/s. We therefore accept that $\sigma = 15$ ft/s. Note in this example that our interest centered around whether the unknown $\sigma = 15$ ft/s, so we used the two-sided or two-tailed test. Had we raised the question concerning whether σ were as large as, say, 20 ft/s, or perhaps as low as, say, 10 ft/s, the upper or lower percentage point, respectively, would have been used.

4-4.4 CONFIDENCE BOUNDS ON THE UNKNOWN POPULATION VARIANCE OR STANDARD DEVIATION

Clearly the chance that chi-square will lie between the lower and upper $\alpha(< 0.5)$ probability levels of its distribution for ν df is

$$Pr[\chi^2_{\alpha}(\nu) \leq \Sigma(x_i - \bar{x})^2/\sigma^2 \leq \chi^2_{1-\alpha}(\nu)] = 1 - 2\alpha \quad (4\text{-}63)$$

TABLE 4-5

PERCENTILES OF THE χ^2 DISTRIBUTION (Ref. 5)



Values of χ^2_P corresponding to P

ν df	$\chi^2_{0.005}$	$\chi^2_{0.01}$	$\chi^2_{0.025}$	$\chi^2_{0.05}$	$\chi^2_{0.10}$	$\chi^2_{0.90}$	$\chi^2_{0.95}$	$\chi^2_{0.975}$	$\chi^2_{0.99}$	$\chi^2_{0.995}$
1	0.000039	0.00016	0.00098	0.0039	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.1026	0.2107	4.61	5.99	7.38	9.21	10.60
3	0.0717	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
120	83.85	86.92	91.58	95.70	100.62	140.23	146.57	152.21	158.95	163.64

For large degrees of freedom

$$\chi^2_p \approx (1/2)\,(z_p + \sqrt{2\nu - 1})^2$$

where

ν = df

z_P = a unit normal variate at probability p

From *Introduction to Statistical Analysis* by W. J. Dixon and F. J. Massey. Copyright©1957 by McGraw-Hill Book Company. Used by permission of McGraw-Hill Book Company.

where

α = probability level

χ^2_α = αth probability level or percentage point of chi-square

$\chi^2_{1-\alpha}$ = $(1-\alpha)$th probability level or percentage point of chi-square.

This confidence statement may easily be transformed to obtain a $(1-2\alpha)$ confidence bound on or about σ^2 or σ, i.e.,

$$Pr[\Sigma(x_i - \bar{x})^2/\chi^2_{1-\alpha} \leq \sigma^2 \leq \Sigma(x_i - \bar{x})^2/\chi^2_{\alpha}]$$
$$= Pr\{[\Sigma(x_i - \bar{x})^2]^{1/2}/\chi_{1-\alpha} \leq \sigma \leq [\Sigma(x_i - \bar{x})^2]^{1/2}/\chi_{\alpha}\} = 1 - 2\alpha. \quad (4\text{-}64)$$

The upper and lower confidence bounds of Eq. 4-64 are called the equal-tail confidence bounds for σ^2 or σ, respectively. The equal-tail confidence bounds in Eq. 4-64, however, do not give the shortest confidence interval about the unknown σ^2 or σ. In fact, to obtain the shortest confidence interval on σ^2, at confidence level $(1-\alpha)$, instead of using $\chi^2_{\alpha/2}$ and $\chi^2_{(1-\alpha/2)}$ as divisors of the SS about the sample mean, one must find numbers $\chi^2_a < \chi^2_b$ such that the length L of the obtained confidence interval given generally by

$$L = \Sigma(x_i - \bar{x})^2\left(\frac{1}{\chi^2_a} - \frac{1}{\chi^2_b}\right) \quad (4\text{-}65)$$

is a minimum, subject to, the degree of confidence $(1-\alpha)$ obtained by the condition

$$\int_{\chi^2_a}^{\chi^2_b} f(\chi^2)\, d\chi^2 = 1 - \alpha \quad (4\text{-}66)$$

where

χ^2_a = lower limit of chi-squared distribution

χ^2_b = upper limit of chi-squared distribution.

The minimum length confidence bounds for σ^2 have been calculated in accordance with Eqs. 4-65 and 4-66 by Tate and Klett (Ref. 14), and their bounds are given in Table 4-6.

It is of interest at this point to cite a comparison of the differences in (relative) lengths of confidence bounds for the equal-tail interval as compared to that of the minimum length interval. For example, refer to Table 4-5 for $\nu = 5$ df and the 0.005 and 0.995 probability levels, which amount to a confidence level of 99%. Here we see that $1/0.412 - 1/16.75 = 2.367$, ignoring for the moment the sum of squares about the sample mean. On the other hand, if we refer to Table 4-6 for the minimum length 99% confidence levels for $\nu = 5$, we have, for the similar calculation, that $1/0.5534 - 1/28.0269 = 1.771$. Thus the difference is of practical significance and would magnify considerably for relatively large sums of squares. It can be expected, therefore, that for unsymmetrical distributions there will clearly be some important differences between the equal-tail area confidence bounds and those of minimum length. On the other hand, one can actually find some cases where the equal-tail area bounds are nearly the same as the minimum bounds in length. For example, consider a comparison of the 99% confidence bounds for $\nu = 24$ df. In this case for the equal-tail area confidence bounds, we have the relative length (ignoring SS) of $1/9.89 - 1/45.56 = 0.079$ from Table 4-5, whereas for the 99% minimum length bounds from Table 4-6, we get $1/10.7169 - 1/51.5619 = 0.074$, or equal intervals. One would expect, of course, that for a large number of df ν the lengths become equivalent due to symmetry.

Another method for determining a confidence interval on the normal population variance is that due to Neyman (Ref. 15). If we use a confidence level of $(1-\alpha)$, say, and consider intervals that cover some hypothesized value, call it σ_0^2 of σ^2, then Neyman (Ref. 15) defines that interval I to be unbiased if

$$Pr\,[(I \text{ covers } \sigma_0^2)|\mu, \sigma^2] \geq 1 - \alpha \text{ if } \sigma_0 = \sigma \quad (4\text{-}67)$$

and

$$Pr\,[(I \text{ covers } \sigma_0^2)|\mu, \sigma^2] < 1 - \alpha \text{ if } \sigma_0 \neq \sigma. \quad (4\text{-}68)$$

(In other words, the chance of coverage has a maximum when $\sigma_0 = \sigma$.) Then the shortest unbiased Neyman interval, which is labeled as I_{SU}, is that interval which satisfies Eqs. 4-67 and 4-68 and for which the left member of Eq. 4-68 is a minimum uniformly for all values of μ, σ^2, and σ_0^2. Tate and Klett (Ref. 14) have also calculated the shortest unbiased Neyman confidence intervals for the normal population variance, and we give their tables as Table 4-7.

TABLE 4-6

DIVISORS FOR THE CONFIDENCE INTERVAL ABOUT NORMAL POPULATION VARIANCE OF MINIMUM LENGTH (Ref. 14)

$$I_{ML} = [\Sigma(x_i - \bar{x})^2 / \chi_b^2,\ \Sigma(x_i - \bar{x})^2 / \chi_a^2]^*$$

Confidence Coefficient $(1 - \alpha)$, $\nu = (n - 1)$ df usually

χ_a^2 = lower limit of chi-squared distribution and

χ_b^2 = upper limit of chi-squared distribution.

ν \ $1 - \alpha$	0.900	0.950	0.990	0.995	0.999
2	0.2104 18.0077	0.1025 21.4812	0.0201 29.1362	0.0100 32.3240	0.0020 39.5708
3	0.5821 17.6381	0.3513 20.7437	0.1148 27.5102	0.0717 30.3027	0.0244 36.5959
4	1.0561 18.1062	0.7082 21.0632	0.2969 27.4603	0.2069 30.0848	0.0908 35.9845
5	1.5938 18.9081	1.1392 21.8001	0.5534 28.0269	0.4113 30.5697	0.2102 36.2654
6	2.1750 19.8739	1.6233 22.7410	0.8700 28.8928	0.6747 31.3966	0.3806 36.9947
7	2.7883 20.9303	2.1473 23.7944	1.2350 29.9229	0.9871 32.4106	0.5979 37.9541
8	3.4262 22.0405	2.7027 24.9147	1.6397 31.0507	1.3406 33.5358	0.8560 39.0631
9	4.0840 23.1844	3.2836 26.0769	2.0775 32.2397	1.7288 34.7308	1.1499 40.2631
10	4.7584 24.3498	3.8855 27.2662	2.5434 33.4685	2.1469 35.9714	1.4755 41.5223
11	5.4467 25.5294	4.5054 28.4733	3.0334 34.7240	2.5906 37.2430	1.8287 42.8238
12	6.1472 26.7180	5.1409 29.6920	3.5447 35.9963	3.0573 38.5330	2.2078 44.1445
13	6.8583 27.9126	5.7899 30.9184	4.0744 37.2809	3.5439 39.8378	2.6086 45.4880
14	7.5788 29.1109	6.4510 32.1497	4.6205 38.5733	4.0483 41.1517	3.0296 46.8441
15	8.3078 30.3113	7.1227 33.3842	5.1813 39.8715	4.5685 42.4732	3.4676 48.2150
16	9.0446 31.5125	7.8043 34.6197	5.7559 41.1710	5.1040 43.7951	3.9248 49.5766
17	9.7883 32.7139	8.4947 35.8560	6.3425 42.4728	5.6523 45.1206	4.3954 50.9511

*I_{ML} = confidence interval of minimum length

(cont'd on next page)

TABLE 4-6 (cont'd)

ν \ $1-\alpha$	0.900	0.950	0.990	0.995	0.999
18	10.5385	9.1932	6.9402	6.2128	4.8806
	33.9148	37.0919	43.7748	46.4465	52.3245
19	11.2947	9.8991	7.5481	6.7846	5.3786
	35.1148	38.3271	45.0765	47.7723	53.6990
20	12.0563	10.6119	8.1654	7.3666	5.8882
	36.3137	39.5611	46.3772	49.0974	55.0743
21	12.8230	11.3310	8.7915	7.9580	6.4085
	37.5112	40.7936	47.6767	50.4216	56.4507
22	13.5946	12.0561	9.4259	8.5588	6.9406
	38.7070	42.0243	48.9736	51.7426	57.8190
23	14.3706	12.7868	10.0679	9.1679	7.4824
	39.9011	43.2532	50.2686	53.0616	59.1857
24	15.1508	13.5227	10.7169	9.7845	8.0322
	41.0935	44.4802	51.5619	54.3793	60.5545
25	15.9351	14.2636	11.3728	10.4088	8.5919
	42.2840	45.7051	52.8521	55.6935	61.9157
26	16.7230	15.0090	12.0348	11.0396	9.1580
	43.4728	46.9281	54.1407	57.0065	63.2808
27	17.5145	15.7587	12.7024	11.6764	9.7293
	44.6598	48.1491	55.4277	58.3186	64.6514
28	18.3095	16.5128	13.3767	12.3211	10.3146
	45.8446	49.3675	56.7096	59.6230	65.9955
29	19.1076	17.2706	14.0554	12.9699	10.9003
	47.0279	50.5843	57.9914	60.9295	67.3589

Reprinted with permission. Copyright © by the American Statistical Association.

In their paper of 1959 Tate and Klett (Ref. 14) raised two questions of interest concerning confidence bounds on the normal population variance:

1. "Does the interval of shortest length based on the sample mean and sample SS depend only on the sample SS?"

2. "Among those intervals based only on the sample SS, is the interval of shortest length necessarily of the form given by the SS divided by two numbers, say a and b, which depend on the sample size n?".

In 1972 these two questions were answered by Cohen (Ref. 16), who determined that the answer to Question No. 1 is "no", for Cohen determined intervals of the proper length whose chance of coverage uniformly in μ and σ was found to be greater than $(1-\alpha)$. However, Cohen (Ref. 16) found that the answer to Question No. 2 is "yes" since he showed that, if one only observes the sample SS about the mean and notes that it divided by the population σ^2 follows the chi-square distribution, there is no other confidence interval with probability of coverage greater than or equal to the confidence level $(1-\alpha)$. Thus it would seem that at least the more important practical questions concerning confidence bounds on any normal population variance are settled.

We conclude our discussion of the normal population variance and its related chi-square distribution with Example 4-4 involving all three types of confidence bounds on the unknown population standard deviation sigma.

TABLE 4-7

DIVISORS FOR NEYMAN'S "SHORTEST" UNBIASED CONFIDENCE INTERVAL FOR NORMAL POPULATION VARIANCE (Ref. 14)

$I_{SU} = [\Sigma(x_i - \bar{x})^2/\chi_b^2,\ \Sigma(x_i - \bar{x})^2/\chi_a^2]$*

Confidence Coefficient $(1 - \alpha)$, $\nu = (n - 1)$ df usually

χ_a^2 = lower limit of chi-squared distribution and
χ_b^2 = upper limit of chi-squared distribution.

ν \ $1-\alpha$	0.900	0.950	0.990	0.995	0.999
2	0.1676 7.8643	0.0847 9.5303	0.0175 13.2855	0.0088 14.8647	0.0018 18.4677
3	0.4764 9.4338	0.2962 11.1915	0.1010 15.1270	0.0639 16.7754	0.0221 20.5244
4	0.8827 10.9583	0.6070 12.8024	0.2640 16.9014	0.1859 18.6106	0.0831 22.4855
5	1.3547 12.4424	0.9892 14.3686	0.4962 18.6214	0.3723 20.3866	0.1933 24.3799
6	1.8746 13.8922	1.4250 15.8966	0.7856 20.2956	0.6144 22.1139	0.3519 26.2160
7	2.4313 15.3136	1.9026 17.3923	1.1221 21.9310	0.9037 23.8001	0.5548 28.0053
8	3.0173 16.7108	2.4139 18.8604	1.4978 23.5328	1.2331 25.4506	0.7972 29.7547
9	3.6276 18.0874	2.9532 20.3050	1.9068 25.1058	1.5969 27.0705	1.0743 31.4709
10	4.2582 19.4463	3.5162 21.7289	2.3444 26.6531	1.9905 28.6628	1.3827 33.1543
11	4.9063 20.7895	4.0995 23.1347	2.8069 28.1779	2.4102 30.2309	1.7188 34.8097
12	5.5696 22.1190	4.7005 24.5247	3.2912 29.6833	2.8528 31.7786	2.0790 36.4463
13	6.2462 23.4362	5.3171 25.9004	3.7948 31.1710	3.3158 33.3080	2.4609 38.0646
14	6.9347 24.7423	5.9477 27.2630	4.3161 32.6414	3.7979 34.8174	2.8650 39.6507
15	7.6340 26.0385	6.5909 28.6141	4.8531 34.0970	4.2965 36.3114	3.2872 41.2209
16	8.3427 27.3257	7.2453 29.9546	5.4041 35.5402	4.8100 37.7927	3.7248 42.7826
17	9.0603 28.6047	7.9099 31.2855	5.9681 36.9711	5.3373 39.2609	4.1775 44.3309

*I_{SU} = Neyman's shortest unbiased confidence interval

(cont'd on next page)

TABLE 4-7 (cont'd)

ν \ $1-\alpha$	0.900	0.950	0.990	0.995	0.999
18	9.7859	8.5842	6.5444	5.8780	4.6467
	29.8759	32.6072	38.3896	40.7147	45.8546
19	10.5188	9.2670	7.1314	6.4300	5.1272
	31.1401	33.9209	39.7984	42.1590	47.3738
20	11.2586	9.9579	7.7290	6.9938	5.6218
	32.3978	35.2267	41.1966	43.5912	48.8733
21	12.0046	10.6562	8.3360	7.5671	6.1281
	33.6494	36.5253	42.5856	45.0132	50.3610
22	12.7565	11.3614	8.9515	8.1496	6.6428
	34.8954	37.8176	43.9672	46.4282	51.8481
23	13.5138	12.0730	9.5752	8.7410	7.1671
	36.1362	39.1036	45.3409	47.8348	53.3266
24	14.2764	12.7908	10.2072	9.3416	7.7043
	37.3719	40.3835	46.7057	49.2305	54.7826
25	15.0437	13.5142	10.8462	9.9493	8.2475
	3.6030	41.6581	48.0645	50.620	56.2408
26	15.8155	14.2430	11.4923	10.5649	8.8015
	39.8296	42.9273	49.4157	52.0024	57.6820
27	16.5917	14.9769	12.1447	11.1874	9.3624
	41.0521	44.1916	50.7610	53.3778	59.1196
28	17.3718	15.7155	12.8033	11.8165	9.9309
	42.2706	45.4514	52.1004	54.7466	60.5496
29	18.1558	16.4586	13.4674	12.4511	10.5035
	43.4855	46.7069	53.4350	56.1114	61.9829

Reprinted with permission. Copyright © by the American Statistical Association.

Example 4-4:

Use the data of Example 4-1 to determine and to compare the lengths of the 95% confidence bounds on the unknown population standard deviation σ for (1) the equal tails case, (2) the minimum length confidence bounds, and (3) the Neyman shortest unbiased confidence bounds.

For the 95% confidence bounds, we find, from Table 4-5, for $\nu = 10$ df and the second form of Eq. 4-62 that

$$Pr\left[\sqrt{1050.55/20.48} \leq \sigma \leq \sqrt{1050.55/3.25}\right]$$

$$= Pr\,[7.16 \leq \sigma \text{ (equal tails)} \leq 17.98] = 0.95$$

the length of which is

$$17.98 - 7.16 = 10.82 \text{ ft/s}.$$

For the minimum length 95% confidence bounds about σ, we determine with the aid of Table 4-6 that

$$Pr\left[\sqrt{1050.55/27.2662} \leq \sigma_{ML} \leq \sqrt{1050.55/3.8855}\right]$$

$$= Pr[6.21 \leq \sigma_{ML} \leq 16.44] = 0.95$$

the length of which is

$$16.44 - 6.21 = 10.23 \text{ ft/s}.$$

For the shortest unbiased Neyman 95% confidence bounds, we see using Table 4-7 that

$$Pr\left[\sqrt{1050.55/21.7289} \leq \sigma_{SU} \leq \sqrt{1050.55/3.5162}\right]$$

$$= Pr[6.95 \leq \sigma_{SU} \leq 17.29] = 0.95$$

the length of which is

$$17.29 - 6.95 = 10.34 \text{ ft/s}.$$

Finally, we note that although there is not a great deal of difference in confidence bound lengths, the end points are nevertheless shifted.

4-4.5 THE APPROXIMATE CHI-SQUARE DISTRIBUTION

There are a rather large number of distributional problems in many Army applications for which one can find a chi-square type of approximation or fit. The "approximate chi-square" involves a two-moment approximation, i.e., the use of the mean and the variance of the statistic of interest. Generally speaking, the approximate chi-square involves the fitting of a new random variable to a quadratic form, of which we desire the probability distribution, or an approximation of some other distribution that is sufficiently accurate for practical applications. It is easy to apply the suggested technique, of which we will give only a schematic view since the principles are thoroughly covered in Ref. 17.

Quite generally, we may deal with two (or more) random variables x and y, which are normally distributed with even different means and variances, and consider a quadratic form $Q = Q(x, y)$ of the variates. Since for normally distributed variables x and y we can find means and variances individually, it is often easy to find the mean m and variance v of the quadratic form Q. Thus in a straightforward manner we have that

$$E(Q) = E[Q(x,y)] = m \tag{4-69}$$

and

$$\text{Var}(Q) = \text{Var}[Q(x,y)] = v. \tag{4-70}$$

Then it can be shown (Ref. 17) that to a good approximation

$$2mQ/v \approx \chi^2(2m^2/v) \tag{4-71}$$

or that the random variable $2mQ/v$ is approximately distributed as χ^2 with $2m^2/v$ df. One can see that some difficulty may be involved in using the chi-square approximation (Eq. 4-71) because the number of df $2m^2/v$ will usually be fractional. However, this problem can always be circumvented by using the Wilson-Hilferty transformation (Ref. 18) of chi-square to a normal variate.

We will illustrate the approximate chi-square technique briefly by using the sample variance s^2, i.e., the quadratic form of Eq. 4-2, which should reproduce χ^2 exactly with $(n - 1)$ df if the approximation has merit. In this case, as previously indicated,

$$m = E(s^2) = \sigma^2$$
$$v = \mathrm{Var}(s^2) = 2\sigma^4/(n-1).$$

Thus

$$2mQ/v = 2(\sigma^2)s^2/[2\sigma^4/(n-1)]$$
$$= \chi^2(2m^2/v) = \chi^2\{2\sigma_4/[2\sigma_4/(n-1)]\} = \chi^2(n-1)$$

precisely with $(n-1)$ df as it should.

Many applications of the approximate chi-square and its accuracy are given in connection with the probability of hitting problems in Ref. 17 and Chapters 14 and 20 of Ref. 19. Moreover, excellent use of the technique extends easily to confidence bounds on system reliability as covered in Chapter 21 of Ref. 19 also.

In terms of the mean m and variance v of the quadratic form $Q(x,y)$ and the fractional number of df, the Wilson-Hilferty transformation t (Ref. 18) becomes

$$t \approx \{3Q^{1/3}m^{2/3} - [3m - v/(3m)]\}/\sqrt{v} \tag{4-72}$$

where t is approximately $N(0,1)$, i.e., normally distributed with mean zero and sigma equal to unity.

4-5 THE SNEDECOR-FISHER VARIANCE RATIO OR F DISTRIBUTION

While the chi-square distribution of Eq. 4-41 is very useful in determining confidence bounds on the unknown normal population variance or sigma—as in Eqs. 4-64, 4-65, and 4-66—it is not very often that we have a sufficiently large sample to estimate the population variance or sigma with great precision. In fact, we are often interested in testing a new product, type of ammunition, or new weapon against an old one, or equivalently in "comparing two normal populaions" sampled for the purpose. It is frequently for such reasons that the statistician is faced with the problem of determining whether two unknown normal population standard deviations are equal on the basis of relatively small samples drawn therefrom. This type of comparison is made possible through the use of the well-known Snedecor-Fisher variance ratio statistic, or, as it is often called, the Snedecor F test.

First, we consider two distinct normal populations that generally may have unknown true means and unknown population standard deviations or variances. Thus we have, quite generally, one normal population with unknown mean μ_1 and standard deviation σ_1, designated by $N(\mu_1,\sigma_1)$ and another one with unknown mean and standard deviation given by μ_2 and σ_2 and designated by $N(\mu_2,\sigma_2)$. In practice, we draw a sample of size n_1 from the first normal population and a sample of size n_2 from the second one. This leads to two sample variances—one from each of the two normal populations—which we will designate by s_1^2 and s_2^2 as follows:

$$s_1^2 = \sum_{i=1}^{n_1}(x_{1i} - \bar{x}_1)^2/(n_1 - 1) \tag{4-73}$$

and

$$s_2^2 = \sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2/(n_2 - 1) \tag{4-74}$$

where

$\bar{x}_1$ = sample mean of first sample
$\bar{x}_2$ = sample mean of second sample.

The Snedecor-Fisher F ratio for testing equality of sigmas, i.e., $\sigma_1 = \sigma_2$, is simply

$$F = s_1^2/s_2^2. \tag{4-75}$$

Quite generally, however, if we have two independent chi-squares, or χ_1^2 with ν_1 df and χ_2^2 with ν_2 df, the ratio

$$F = (\chi_1^2/\nu_1)/(\chi_2^2/\nu_2) \tag{4-76}$$

follows the Snedecor F distribution with ν_1 and ν_2 df, respectively. Note that ν_1 is taken as the numerator number of df.

The pdf of the random variable F is given by

$$f(F) = \frac{\Gamma(\nu_1/2 + \nu_2/2)\,(\nu_1/\nu_2)^{\nu_1/2} F^{(\nu_1/2)-1}}{\Gamma(\nu_1/2)\,\Gamma(\nu_2/2)\,[1 + \nu_1 F/\nu_2]^{(\nu_1+\nu_2)/2}}. \tag{4-77}$$

The rth moment μ_r' of F about the origin is easily found (by taking the ratio of moments of the two independent chi-squares) to be

$$\mu_r' = \mu_r'(F) = (\nu_2/\nu_1)^r \Gamma(r + \nu_1/2)\,\Gamma(-r + \nu_2/2)/\Gamma(\nu_1/2)\,\Gamma(\nu_2/2). \tag{4-78}$$

The mean value of F depends only on the denominator df and is

$$E(F) = \nu_2/(\nu_2 - 2),\ \nu_2 > 2 \tag{4-79}$$

which clearly approaches unity for large ν_2.

The variance of the statistic F is

$$\mathrm{Var}(F) = \sigma^2(F) = 2\nu_2^2\,(\nu_1 + \nu_2 - 2)/[\nu_1(\nu_2 - 2)^2(\nu_2 - 4)],\ \nu_2 > 4. \tag{4-80}$$

Whereas the mean of F approaches the limit of unity for large ν_2, the variance of the F for large ν_1 and ν_2 does indeed approach zero as can be seen from Eq. 4-80.

The skewness α_3 and kurtosis α_4 coefficients for F are rather complicated, i.e.,

$$\alpha_3 = \alpha_3(F) = \sqrt{\beta_1} = \frac{\sqrt{8(\nu_2 - 4)}\,(2\nu_1 + \nu_2 - 2)}{\sqrt{(\nu_1 + \nu_2 - 2)\nu_1}\,(\nu_2 - 6)} \tag{4-81}$$

$$\alpha_4 = \alpha_4(F) = 3 + \frac{12[(\nu_2 - 2)^2(\nu_2 - 4) + \nu_1(\nu_1 + \nu_2 - 2)\,(5\nu_2 - 22)]}{\nu_1\,(\nu_2 - 6)\,(\nu_2 - 8)\,(\nu_1 + \nu_2 - 2)} \tag{4-82}$$

Note that for large numbers of df ν_1 and ν_2 the skewness does approach zero and α_4 approaches 3 as for the normal distribution.

The Snedecor F is related to R. A. Fisher's z by the equality

$$z = (1/2)\ln F. \tag{4-83}$$

Also one may note from Eq. 4-77 that there is a definite relation between the random variable F and Karl Pearson's incomplete beta function (Ref. 20). In fact, if x is a beta variate and x_α is the α probability level or percentage point, then

$$Pr[F > F_\alpha] = I_{x_\alpha}(\nu_1/2,\ \nu_2/2) \tag{4-84}$$

where F_α is the α probability level of F and

$$F_\alpha = \nu_2 x_\alpha/[\nu_1(1 - x_\alpha)] \tag{4-85}$$

(The right-hand side (RHS) of Eq. 4-84 is Karl Pearson's incomplete beta function (Ref. 20).)

The 90%, 95%, 97.5%, and 99%, or probability, levels of $F(\nu_1,\nu_2)$ are given in Table 4-8 and were reproduced from Ref. 5. If we designate $F_{1-\alpha}(\nu_1,\nu_2)$ as the upper α significance level, the lower probability levels are found from

$$F_{\alpha}(\nu_1,\nu_2) = 1/F_{1-\alpha}(\nu_2,\nu_1). \tag{4-86}$$

Example 4-5:

Some difficulty was being experienced with the MV dispersion of a 20-mm high-velocity projectile. In fact, for firings at a vertical target the relatively large dispersion in the vertical direction was attributable to MV dispersion. A new propellant was developed and rotating bands were applied more uniformly with the result that the designers indicated the bivariate dispersion pattern should be "absolutely circular". Ten sample rounds of the new or improved 20-mm projectile were fired for impact on a vertical target placed at 200 m. The horizontal impact points from the left-most round and vertical impacts from the bottom round measured in inches are given in Table 4-9.

Is there any statistical evidence that $\sigma_x \neq \sigma_y$?

After identifying the horizontal impact points as x and the vertical ones as y, we calculate, by Eqs. 4-73 and 4-74, with $\nu = 10 - 1 = 9$ df

$$s_x = 14.73,\ s_y = 14.95.$$

Hence, by Eq. 4-75,

$$F = s_x^2/s_y^2 = 0.97$$

which for $\nu_1 = \nu_2 = 9$ df referred to Table 4-8 is not statistically significant even at the 80% level since $F_{0.90}(9,9) = 2.44$ and $F_{0.10}(9,9) = 1/2.44 = 0.41$. (Note that we are using a two-tailed test.) We conclude, therefore, that the improved projectile may indeed exhibit circularity for its dispersion pattern. Moreover, for the purpose of weapon systems analyses, one may treat the 20-mm weapon-ammunition combination as having a circular normal distribution of delivery errors with the "circular" sigma $\hat{\sigma}$ at 200 m given by

$$\sigma = \sqrt{[(14.73)^2 + (14.95)^2]/2} = 14.84 \text{ in.}$$

which may be converted to equivalent angular mils.

As a final comment on approximations, the Fisher z of Eq. 4-83 is more nearly normally distributed than is the F statistic of Eq. 4-75. The Wilson-Hilferty approximation (Ref. 18), or cube-root transformation of F, can be used to obtain a variate, call it z, which is, for practical purposes, distributed as a unit normal variate. This technique involves putting

$$z = \{[1 - 2/(9\nu_2)]F^{1/3} - [1 - 2/(9\nu_1)]\}\,[2F^{2/3}/(9\nu_2) + 2/(9\nu_1)]^{-1/2} \tag{4-87}$$

where z is approximately normally distributed, i.e.,

$$z \approx N(0,1).$$

The values of z are easily calculated by Eq. 4-87 with a scientific-type pocket calculator for reference to normal tables.

We will discuss the problem of comparing more than two variances next.

TABLE 4-8

PERCENTILES OF THE F DISTRIBUTION (Ref. 5)



$F_{0.90}(\nu_1,\nu_2)$

ν_1 = degrees of freedom for numerator

ν_2 = degrees of freedom for denominator

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	60.71	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.20	5.18	5.18	5.17	5.16	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.14	3.12	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.50	2.50	2.46	2.42	2.40	2.38	2.36	2.34	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.38	2.34	2.30	2.28	2.25	2.23	2.21	2.18	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.28	2.24	2.20	2.18	2.16	2.13	2.11	2.08	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.03	2.00	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.15	2.10	2.06	2.04	2.01	1.99	1.96	1.93	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.90	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.05	2.01	1.96	1.94	1.91	1.89	1.86	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.02	1.97	1.92	1.90	1.87	1.85	1.82	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.99	1.94	1.89	1.87	1.84	1.81	1.78	1.75	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.96	1.91	1.86	1.84	1.81	1.78	1.75	1.72	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.93	1.89	1.84	1.81	1.78	1.75	1.72	1.69	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.91	1.86	1.81	1.79	1.76	1.73	1.70	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.89	1.84	1.79	1.77	1.74	1.71	1.68	1.64	1.61
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.87	1.83	1.78	1.75	1.72	1.69	1.66	1.62	1.59
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.86	1.81	1.76	1.73	1.70	1.67	1.64	1.60	1.57
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.84	1.80	1.74	1.72	1.69	1.66	1.62	1.59	1.55
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.83	1.78	1.73	1.70	1.67	1.64	1.61	1.57	1.53
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.82	1.77	1.72	1.69	1.66	1.63	1.59	1.56	1.52
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.81	1.76	1.71	1.68	1.65	1.61	1.58	1.54	1.50
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85	1.80	1.75	1.70	1.67	1.64	1.60	1.57	1.53	1.49
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.79	1.74	1.69	1.66	1.63	1.59	1.56	1.52	1.48
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83	1.78	1.73	1.68	1.65	1.62	1.58	1.55	1.51	1.47
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.77	1.72	1.67	1.64	1.61	1.57	1.54	1.50	1.46
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.71	1.66	1.61	1.57	1.54	1.51	1.47	1.42	1.38
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.66	1.60	1.54	1.51	1.48	1.44	1.40	1.35	1.29
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.60	1.55	1.48	1.45	1.41	1.37	1.32	1.26	1.19
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.55	1.49	1.42	1.38	1.34	1.30	1.24	1.17	1.00

(cont'd on next page)

TABLE 4-8 (cont'd)

PERCENTILES OF THE F DISTRIBUTION

$F_{0.95}(\nu_1,\nu_2)$

ν_1 = degrees of freedom for numerator

ν_2 = degrees of freedom for denominator

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

(cont'd on next page)

TABLE 4-8 (cont'd)

PERCENTILES OF THE F DISTRIBUTION

$F_{0.975}(\nu_1, \nu_2)$

ν_1 = degrees of freedom for numerator

ν_2 = degrees of freedom for denominator

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.2	1001	1006	1010	1014	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.28	6.23	6.18	6.12	6.07	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.36	4.31	4.25	4.20	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.61	3.56	3.51	3.45	3.39	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.37	3.31	3.26	3.20	3.14	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.17	3.12	3.06	3.00	2.94	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	3.02	2.96	2.91	2.85	2.79	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.89	2.84	2.78	2.72	2.66	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.79	2.73	2.67	2.61	2.55	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.70	2.64	2.59	2.52	2.46	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.63	2.57	2.51	2.45	2.38	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.56	2.50	2.44	2.38	2.32	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.50	2.44	2.38	2.32	2.26	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.45	2.39	2.33	2.27	2.20	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.41	2.35	2.29	2.22	2.16	2.09
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.64	2.53	2.42	2.37	2.31	2.25	2.18	2.11	2.04
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.60	2.50	2.39	2.33	2.27	2.21	2.14	2.08	2.00
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.57	2.47	2.36	2.30	2.24	2.18	2.11	2.04	1.97
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.54	2.44	2.33	2.27	2.21	2.15	2.08	2.01	1.94
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51	2.41	2.30	2.24	2.18	2.12	2.05	1.98	1.91
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.49	2.39	2.28	2.22	2.16	2.09	2.03	1.95	1.88
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57	2.47	2.36	2.25	2.19	2.13	2.07	2.00	1.93	1.85
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.45	2.34	2.23	2.17	2.11	2.05	1.98	1.91	1.83
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43	2.32	2.21	2.15	2.09	2.03	1.96	1.89	1.81
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.14	2.07	2.01	1.94	1.87	1.79
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.29	2.18	2.07	2.01	1.94	1.88	1.80	1.72	1.64
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06	1.94	1.88	1.82	1.74	1.67	1.58	1.48
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94	1.82	1.76	1.69	1.61	1.53	1.43	1.31
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.64	1.57	1.48	1.39	1.27	1.00

(cont'd on next page)

TABLE 4-8 (cont'd)

PERCENTILES OF THE F DISTRIBUTION

$F_{0.99}(\nu_1, \nu_2)$

ν_1 = degrees of freedom for numerator

ν_2 = degrees of freedom for denominator

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

From *Introduction to Statistical Analysis* by W. J. Dixon and F. J. Massey. Copyright © 1957 by McGraw-Hill Book Company. Used by permission of McGraw-Hill Book Company.

TABLE 4-9. IMPACT POINTS

Horizontal	Vertical
0	18.20
11.47	49.63
13.48	9.94
15.60	36.36
17.27	10.82
17.61	10.60
21.87	0
22.68	25.47
23.67	21.97
57.17	7.20

4-6 SIGNIFICANCE TESTS FOR THE EQUALITY OF SEVERAL POPULATION VARIANCES

4-6.1 PRELIMINARY REMARKS

The problem of comparing or determining whether two population variances can be considered to be equal having been covered, it is a natural extension that one may have at hand several sample variances and may wish to establish whether or not they represent samples from normal populations with equal true variances, i.e., exhibit "homogeneity of variances".* Again, this is done by calculating the value of sample statistics that may be referred to an appropriate table of percentage points of the relevant probability distribution. In other words, it amounts to an extension of the Snedecor-Fisher F statistic. Although for only two observed sample variances it is natural to use the ratio of them, there can be a variety of ways of combining several sample variances in an appropriate significance test. In fact, this is what has occurred over the years, and as a result, there are, as would be expected, several different tests for homogeneity of variances available in the statistical literature. For the purposes of this handbook, we will include Bartlett's statistic (Refs. 21 and 22), Cochran's statistic (Ref. 23), Hartley's maximum variance ratio statistic (Ref. 24), Cadwell's statistic (Ref. 25), and Bartlett and Kendall's statistic (Ref. 21). We will present these in sufficient detail to give the practicing Army analyst some background, will comment on their properties, usefulness, and power, and then will give an example.

In the sequel we will consider p random samples from p possibly "different" normal populations and will let

$\nu_i = n_i - 1$ = number of df for the ith sample

s_i^2 = sample variance for the sample of size n_i from the ith normal population.

4-6.2 BARTLETT'S STATISTIC

Bartlett's test (Refs. 21 and 22) is based on the Neyman-Pearson likelihood ratio and in its χ^2 form uses the statistic

$$F_B = \frac{(\Sigma\nu_i)\ln[\Sigma\nu_i s_i^2/(\Sigma\nu_i)] - \Sigma\nu_i \ln s_i^2}{1 + [\Sigma(1/\nu_i) - 1/\Sigma\nu_i]/[3(p-1)]} = M/C \qquad (4\text{-}88)$$

with p, ν_i, and s_i as previously defined. We have designated the numerator of Eq. 4-88 as M and the denominator as C. The percentage points of M are given in the *Biometrika Tables for Statisticians* (Ref. 26), which most Army statisticians should have readily available. The denominator C of Eq. 4-88 might be regarded as a "correction factor" due to Bartlett (Ref. 21) and is used to transform M such that the ratio

$$F_B = M/C \approx \chi^2(p-1) \qquad (4\text{-}89)$$

*Often called "homoscedasticity".

or F_B is distributed approximately as chi-square with $(p-1)$ df, which should be adequate for many practical problems.

Again observing the numerator M of Eq. 4-88, we can obtain the relation between M and the quantity L^*, which is often defined in the statistical literature as Bartlett's statistic. This relation is

$$M = -(\Sigma\nu_i)\ln L^* \tag{4-90}$$

with

$$L^* = [\prod_{i=1}^{p}(s_i^2)^{\nu_i/\Sigma\nu_i}]/[\sum_{i=1}^{p}(\nu_i s_i^2)/\Sigma\nu_i]. \tag{4-91}$$

It is seen in this connection that L^* is really the ratio of the weighted geometric mean of the sample variances to their weighted arithmetic mean. Glaser (Ref. 27) has calculated the exact critical values for L^*, and we give his improved table of percentage points of L^* here as Table 4-10. The null hypothesis is rejected when the observed value of L^* is less than the table value for a lower tail area.

It might be noted or inferred that Bartlett's F_B or L^* represents very efficient statistics for judging "general homoscedasticity" but would not necessarily detect "outlying" variances.

4-6.3 COCHRAN'S STATISTIC

Cochran's statistic F_C (Ref. 23) or test for homoscedasticity employs the ratio of the maximum sample variance to the total of all of them, or

$$F_C = \max(s_1^2, s_2^2, \ldots, s_p^2)/\sum_{i=1}^{p} s_i^2 \tag{4-92}$$

Thus it is seen that Cochran's statistic would in effect test whether the largest sample variance of several such variances is too large based on the total, or sum, of all the sample variances considered. Tables of critical values or percentage points of Cochran's statistic Eq. 4-92 are given in Ref. 26 and also in Dixon and Massey's book (Ref. 28).

4-6.4 HARTLEY'S STATISTIC

As his test of homoscedasticity, Hartley (Ref. 24) uses the maximum F or maximum variance ratio of the sample variances which is

$$F_H = F_{\max} = \max(s_i^2)/\min(s_i^2). \tag{4-93}$$

We note in this connection that the Hartley statistic is very simple to calculate and is used to determine whether the largest and smallest sample variances are "too far apart". It should be noted that if the maximum s_i^2 and the minimum s_i^2 are too discrepant, either or both could possibly represent different populations with one or more anomalous variances.

The upper 5% probability levels of F_H are given in Ref. 24, and David (Ref. 29) gives further tables and includes the 1% points as well. David's tables are given also in the *Biometrika Tables for Statisticians* (Ref. 26). We give David's tables from Ref. 26 as Table 4-11.

4-6.5 CADWELL'S STATISTIC

Instead of using sample variances to test for homogeneity of population variances, Cadwell (Ref. 25) has developed a test based on the ratio of the maximum to the minimum sample ranges and thereby avoids the calculation of variances or SS about sample means. If we refer to the range of the ith sample as r_i, Cadwell's statistic is

$$\max(r_i)/\min(r_i). \tag{4-94}$$

TABLE 4-10

EXACT BARTLETT CRITICAL VALUES (Ref. 27)

ν	α 0.10	0.05	0.01	ν	α 0.10	0.05	0.01
		$p = 3$				$p = 4$	
4	0.6539	0.5762	0.4304	4	0.6507	0.5850	0.4608
5	0.7163	0.6483	0.5149	5	0.7133	0.6559	0.5431
6	0.7600	0.7000	0.5787	6	0.7572	0.7065	0.6045
7	0.7921	0.7387	0.6282	7	0.7895	0.7444	0.6519
8	0.8168	0.7686	0.6676	8	0.8143	0.7737	0.6893
9	0.8362	0.7924	0.6996	9	0.8340	0.7970	0.7196
10	0.8519	0.8118	0.7260	10	0.8498	0.8160	0.7446
11	0.8649	0.8280	0.7483	11	0.8629	0.8318	0.7655
14	0.8931	0.8632	0.7977	14	0.8914	0.8662	0.8119
19	0.9206	0.8980	0.8476	19	0.9194	0.9003	0.8586
24	0.9369	0.9187	0.8779	24	0.9359	0.9205	0.8868
29	0.9477	0.9325	0.8981	29	0.9468	0.9340	0.9056
49	0.9689	0.9597	0.9387	49	0.9683	0.9606	0.9433
99	0.9845	0.9799	0.9693	99	0.9843	0.9804	0.9717
		$p = 5$				$p = 6$	
4	0.6530	0.5952	0.4850	4	0.6566	0.6045	0.5046
5	0.7151	0.6646	0.5653	5	0.7182	0.6727	0.5832
6	0.7587	0.7142	0.6248	6	0.7612	0.7213	0.6410
7	0.7908	0.7512	0.6704	7	0.7930	0.7574	0.6851
8	0.8154	0.7798	0.7062	8	0.8174	0.7854	0.7197
9	0.8349	0.8025	0.7352	9	0.8367	0.8076	0.7475
10	0.8507	0.8210	0.7590	10	0.8523	0.8257	0.7703
11	0.8637	0.8364	0.7789	11	0.8652	0.8407	0.7894
14	0.8920	0.8699	0.8229	14	0.8932	0.8734	0.8315
19	0.9198	0.9031	0.8671	19	0.9207	0.9057	0.8737
24	0.9362	0.9228	0.8936	24	0.9369	0.9249	0.8990
29	0.9471	0.9358	0.9114	29	0.9476	0.9376	0.9159
49	0.9685	0.9617	0.9468	49	0.9688	0.9628	0.9496
99	0.9843	0.9809	0.9734	99	0.9845	0.9815	0.9748
		$p = 7$				$p = 8$	
4	0.6605	0.6126	0.5207	4	0.6642	0.6197	0.5343
5	0.7214	0.6798	0.5978	5	0.7245	0.6860	0.6100
6	0.7640	0.7275	0.6542	6	0.7667	0.7329	0.6652
7	0.7955	0.7629	0.6970	7	0.7978	0.7677	0.7069
8	0.8196	0.7903	0.7305	8	0.8217	0.7946	0.7395
9	0.8386	0.8121	0.7575	9	0.8405	0.8160	0.7657
10	0.8540	0.8298	0.7795	10	0.8557	0.8333	0.7871
11	0.8668	0.8444	0.7980	11	0.8683	0.8477	0.8050
14	0.8944	0.8764	0.8385	14	0.8957	0.8790	0.8443
19	0.9216	0.9080	0.8791	19	0.9226	0.9100	0.8835
24	0.9377	0.9267	0.9034	24	0.9384	0.9283	0.9069
29	0.9483	0.9391	0.9195	29	0.9489	0.9404	0.9225
49	0.9692	0.9637	0.9518	49	0.9696	0.9645	0.9536
99	0.9847	0.9819	0.9759	99	0.9849	0.9823	0.9769

(cont'd on next page)

TABLE 4-10 (cont'd)

ν	α 0.10	0.05	0.01	ν	α 0.10	0.05	0.01
		$p = 9$				$p = 10$	
4	0.6676	0.6260	0.5458	4	0.6708	0.6315	0.5558
5	0.7274	0.6914	0.6204	5	0.7301	0.6961	0.6293
6	0.7692	0.7376	0.6744	6	0.7716	0.7418	0.6824
7	0.8000	0.7719	0.7153	7	0.8021	0.7757	0.7225
8	0.8236	0.7984	0.7471	8	0.8254	0.8017	0.7536
9	0.8423	0.8194	0.7726	9	0.8439	0.8224	0.7786
10	0.8574	0.8365	0.7935	10	0.8588	0.8392	0.7990
11	0.8698	0.8506	0.8109	11	0.8712	0.8531	0.8160
14	0.8969	0.8814	0.8491	14	0.8980	0.8834	0.8532
19	0.9234	0.9117	0.8871	19	0.9243	0.9132	0.8903
24	0.9391	0.9297	0.9099	24	0.9398	0.9309	0.9124
29	0.9495	0.9416	0.9250	29	0.9500	0.9426	0.9271
49	0.9699	0.9652	0.9551	49	0.9703	0.9658	0.9564
99	0.9851	0.9827	0.9776	99	0.9852	0.9830	0.9783

NOTE: p = number of populations; ν = number of degrees of freedom for each sample; α = level of test.

Reprinted with permission. Copyright © by the American Statistical Association.

4-6.6 BARTLETT AND KENDALL'S STATISTIC

Although the analysis of variance (ANOVA) tests usually apply to one-way, two-way, etc., classifications of means, it is often of interest in practice to conduct an ANOVA of sample variances from different sources to determine whether the assumption of homoscedasticity is justified. This type of problem leads to the concept of the Bartlett- and Kendall-type statistic for testing the equality of variances which involves the ANOVA of the logarithms of sample variances. In fact, the logarithms of sample variances for suitably large df approach the normal distribution. The Bartlett-Kendall statistic (Ref. 21) is often referred to as "Log ANOVA" and is computed as follows:

Consider $i = 1, 2, \ldots, p$ possible sources of variation, or possibly "different" normal populations, from which we have several, or m_i, sample variances from the ith population where $m_i > 1$ for at least one of the populations. Then let

$$z_{ij} = \ln s_{ij}^2 = \text{logarithm of } j\text{th sample variance from } i\text{th population} \tag{4-95}$$

$$z_{i.} = \sum_{j=1}^{m_i} (\ln s_{ij}^2)/m_i \tag{4-96}$$

$$z_{..} = \sum_{i=1}^{p} \sum_{j=1}^{m_i} (\ln s_{ij}^2)/(p \sum_{i=1}^{p} m_i) \tag{4-97}$$

where

$z_{i.}$ = ith average of z_{ij}'s

$z_{..}$ = grand average of z_{ij}'s.

Thus the reader may liken our outline to a one-way classification in the ANOVA in which there are at least two observations per cell, and an observation is $\ln s_{ij}^2$. Finally, Bartlett and Kendall's Log ANOVA is calculated as

TABLE 4-11

HARTLEY'S STATISTIC
PERCENTAGE POINTS OF THE RATIO, s^2_{max}/s^2_{min} (Ref. 27)

Upper 5% Points

ν \ p	2	3	4	5	6	7	8	9	10	11	12
2	39·0	87·5	142	202	266	333	403	475	550	626	704
3	15·4	27·8	39·2	50·7	62·0	72·9	83·5	93·9	104	114	124
4	9·60	15·5	20·6	25·2	29·5	33·6	37·5	41·1	44·6	48·0	51·4
5	7·15	10·8	13·7	16·3	18·7	20·8	22·9	24·7	26·5	28·2	29·9
6	5·82	8·38	10·4	12·1	13·7	15·0	16·3	17·5	18·6	19·7	20·7
7	4·99	6·94	8·44	9·70	10·8	11·8	12·7	13·5	14·3	15·1	15·8
8	4·43	6·00	7·18	8·12	9·03	9·78	10·5	11·1	11·7	12·2	12·7
9	4·03	5·34	6·31	7·11	7·80	8·41	8·95	9·45	9·91	10·3	10·7
10	3·72	4·85	5·67	6·34	6·92	7·42	7·87	8·28	8·66	9·01	9·34
12	3·28	4·16	4·79	5·30	5·72	6·09	6·42	6·72	7·00	7·25	7·48
15	2·86	3·54	4·01	4·37	4·68	4·95	5·19	5·40	5·59	5·77	5·93
20	2·46	2·95	3·29	3·54	3·76	3·94	4·10	4·24	4·37	4·49	4·59
30	2·07	2·40	2·61	2·78	2·91	3·02	3·12	3·21	3·29	3·36	3·39
60	1·67	1·85	1·96	2·04	2·11	2·17	2·22	2·26	2·30	2·33	2·36
∞	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00	1·00

Upper 1% Points

ν \ p	2	3	4	5	6	7	8	9	10	11	12
2	199	448	729	1036	1362	1705	2063	2432	2813	3204	3605
3	47·5	85	120	151	184	21(6)	24(9)	28(1)	31(0)	33(7)	36(1)
4	23·2	37	49	59	69	79	89	97	106	113	120
5	14·9	22	28	33	38	42	46	50	54	57	60
6	11·1	15·5	19·1	22	25	27	30	32	34	36	37
7	8·89	12·1	14·5	16·5	18·4	20	22	23	24	26	27
8	7·50	9·9	11·7	13·2	14·5	15·8	16·9	17·9	18·9	19·8	21
9	6·54	8·5	9·9	11·1	12·1	13·1	13·9	14·7	15·3	16·0	16·6
10	5·85	7·4	8·6	9·6	10·4	11·1	11·8	12·4	12·9	13·4	13·9
12	4·91	6·1	6·9	7·6	8·2	8·7	9·1	9·5	9·9	10·2	10·6
15	4·07	4·9	5·5	6·0	6·4	6·7	7·1	7·3	7·5	7·8	8·0
20	3·32	3·8	4·3	4·6	4·9	5·1	5·3	5·5	5·6	5·8	5·9
30	2·63	3·0	3·3	3·4	3·6	3·7	3·8	3·9	4·0	4·1	4·2
60	1·96	2·2	2·3	2·4	2·4	2·5	2·5	2·6	2·6	2·7	2·7
∞	1·00	1·0	1·0	1·0	1·0	1·0	1·0	1·0	1·0	1·0	1·0

s^2_{max} is the largest and s^2_{min} the smallest in a set of p independent mean squares, each based on ν degrees of freedom.

Values in the column $p = 2$ and in the rows $\nu = 2$ and ∞ are exact. Elsewhere the third digit may be in error by a few units for the 5% points and several units for the 1% points. The third digit figures in brackets for $\nu = 3$ are the most uncertain.

Reprinted with permission. Copyright © by Biometrika Trustees.

$$F_{BK} = [\sum_{i=1}^{p}(z_{i\cdot} - z_{\cdot\cdot})^2][\sum_{j=1}^{p}(m_i - 1)] \div \{[\sum_{i=1}^{p}\sum_{j=1}^{m_i}(z_{ij} - z_{i\cdot})]^2(p-1)\} \quad (4\text{-}98)$$

Under the ANOVA assumptions F_{BK} is distributed in probability as the Snedecor-Fisher F statistic.

4-6.7 COMPARISONS OF THE TESTS OF HOMOSCEDASTICITY

Gartside (Ref. 30) has studied the relative effectiveness of all of the previously discussed statistics for judging homoscedasticity. The Gartside study was performed under the assumption of a null hypothesis in which all of the normal population variances are equal, and there are three alternatives—one case for equal sample sizes of $n = 16$ with $(p-1)$ of the population variances $c \neq 1$ times the other population variance, a second case for equal sample sizes and $(p-1)$ of the population variances equal with the last one $c \neq 1$ times as large, and a third in which the second case is repeated but for different sample sizes to study possible effects. Finally, Gartside (Ref. 30) considered sampling a Weibull distribution with shape parameter equal to 4/3, whereas the Weibull universe is approximately normal for a shape parameter of about 10/3. For the Weibull sampling study samples of sizes 4 and 16 were used in this simulation. Gartside was particularly interested in each of the statistics insofar as controlling the Type I error rates of 0.05 and 0.01 were concerned and in the power of the tests to reject the erroneous null hypothesis when the alternative hypothesis was true. As a result of his study, Gartside (Ref. 30) found that Bartlett's statistic was very powerful in all of the experimental situations considered in the study and had good control of the Type I error rates as well. Under the condition of nonnormality, i.e., the Weibull assumption, the only statistic to maintain stable error rates turned out to be the Log ANOVA, or logarithmic transformation, with the ANOVA technique. As is so often true, this further substantiates the "robustness" of the ANOVA-type test even for transformed data involving variances.

Gartside concluded that when the alternative hypothesis is not known (which is certainly the usual situation) and the assumption of normality for the null hypothesis can be relied upon, Bartlett's test would be the best to use. On the other hand, if it is suspected that just one population variance is really larger than the rest, Cochran's test would be a good choice since it maintains power quite well. If a shortcut-type test were necessary, Hartley's and Cadwell's statistics would both perform suitably. Gartside also pointed out that Bartlett's statistic, modified to use the sample range instead of the variance, would be rather good, especially since its power is superior to that of Cadwell's statistic. In fact, we conjecture that the approximate chi-square technique of par. 4-4.5 could be used quite effectively to obtain the approximate number of degrees of freedom for the range, or the square of the range, in Bartlett's type of weighted statistic, for example. Finally, if there are reasons to believe that one is dealing with nonnormal data, the more conservative Log ANOVA approach should probably be used if possible.

4-6.8 FURTHER STUDIES ON HOMOSCEDASTICITY

Beckman and Tietjen (Ref. 31) have developed tables of the upper 10% and 25% points or probability levels of Hartley's maximum F, should one have use of such values. Chambers (Ref. 32) gives an extension of tables of percentage points of Hartley's largest variance ratio for the 0.01 and 0.05 levels and for $p = 6,8,10,11$ (1)15(5)30 with $\nu = 10, 12, 15, 20, 30, 60, \infty$.

For equal sample sizes also Harsaae (Ref. 33) gives tables of percentage points of Bartlett's M for $\alpha = 0.001$, 0.01, 0.05, 0.10; $\nu = 1(1)10$; and $p = 3(1)12$.

Regarding large sample results, Somerville (Ref. 34) discusses the problem of the optimum (minimum) sample size for choosing the population having the smallest variance. Saxena (Ref. 35) presents a study of the problem of interval estimation of the largest variance of several normal populations.

Guenther (Ref. 36) gives some useful techniques for the calculation of factors for tests and determination of confidence intervals concerning the ratio of only two normal population variances, and John (Ref. 37) combines the similar problem of and gives tables for comparing two normal population variances or two gamma distributed means.

Samiuddin and Atiqullah (Ref. 38) use the Wilson-Hilferty cube-root transformation of variances to approximate normality to determine the equality of several variances.

In connection with multiple comparison tests, Tietjen and Beckman (Ref. 39) gives additional tables concerning the application and use of the Hartley type maximum F ratio.

If one has interest in "robust", large-sample tests of homoscedasticity, he should study Layard's paper (Ref. 40) in some detail.

Finally, a study of optimum subsample sizes for the Bartlett-Kendall statistic has been conducted by Toothaker, Hicks, and Price (Ref. 41).

To illustrate the multiple-variance testing technique, we will present the comparison of several normal population variances in Example 4-6.

Example 4-6:

In a development test of a new type of hand grenade, it was claimed that the new grenade could be thrown with improved and especially consistent dispersion in the range direction. Therefore, five infantrymen who had experience in throwing hand grenades were each assigned 15 of the new grenades at random from 75 made up for the purpose, and each of the infantrymen threw his 15 grenades at a stake placed about 30 m from the throwing position. The deviations from the stake in the range and deflection directions were measured, and all of the five sample variances (in ft^2) calculated, based on 14 df. Is there any evidence that homoscedasticity does not exist for the sample variances given in Table 4-12?

TABLE 4-12. SAMPLE VARIANCES

Thrower	Variance in range, ft^2
1	125.29
2	71.16
3	59.67
4	89.17
5	32.42

As a quick test, we could use Hartley's maximum F statistic from Eq. 4-93 to obtain

$$F_H = F_{\max} = 125.29/32.42 = 3.86$$

with $\nu = 14$ df. We see from Table 4-11 that for $\nu = 15$ df and $p = 5$, the upper 5% point of $F_{\max}$ is 4.37. Hence we conclude that the five populations variances are equal.

As a further check with a more powerful test, we will use Bartlett's L^* of Eq. 4-91. Here we see for the equal sample sizes of $n = 15$ that

$$\nu_i = 14,\ \Sigma\nu_i = 70,\ \nu_i/\Sigma\nu_i = 0.2.$$

The calculation of L^* gives

$$L^* = 68.77/75.54 = 0.91.$$

From Table 4-10 of exact Bartlett critical values for $p = 5$ and $\nu = 14$, the 0.91 exceeds the 10% point value of 0.89, so we conclude that homoscedasticity does indeed hold. Hence we may as well use the average variance of the five throwers as the estimate of the population value.

Homoscedasticity is most often a prerequisite to conducting a significance test or ANOVA concerning the equality of population means, especially since the problem of trying to judge the equality of normal populations is conducted on a parameter-by-parameter basis. Therefore, with the preceding treatment of homoscedasticity, we are now ready to proceed with Student's t statistic and its properties and uses.

4-7 STUDENT'S t DISTRIBUTION

4-7.1 INTRODUCTION

One of the striking and important developments in the theory of mathematical statistics concerning the likelihood of occurrence for a sample of size n from a normal population is that the data may be transformed into two distributions—one involving the sample mean that uses a single df and the other the distribution of the SS about the sample mean, or the sample variance, which uses a chi-square distribution with the remaining $(n - 1)$ df of the original sample. Moreover, this leads immediately to Student's t distribution, which is completely free of any population nuisance parameters because the resulting Student's t depends in probability only on the number of df in the sample variance or, that is to say, $(n - 1)$. We may summarize the most useful points by considering a sample of size n from a normal population with mean equal to μ and variance σ^2, or standard deviation σ—i.e., the sample is from $N(\mu, \sigma^2)$. Then if we define the quantity t as

$$t = (\bar{x} - \mu)\sqrt{n}/s \tag{4-99}$$

we have that the pdf of Student's t is

$$f(t) = \frac{[(n-2)/2]!}{[(n-3)/2]!\,\pi\sqrt{n-1}\,\{1 + [t^2/(n-1)]\}^{n/2}} \tag{4-100}$$

$$= \{1/[\sqrt{\nu}\,\beta(1/2, \nu/2)]\}\,[1 + (t^2/\nu)]^{-(\nu+1)/2}$$

where

$\nu = n - 1$ df. [Note: $(1/2)! = \sqrt{\pi}/2$.]

Student's t distribution (Eq. 4-100) is symmetric about the origin as the mean, and hence all odd moments are equal to zero. If we put $r = 2, 4, \ldots$, i.e., an even number, then the rth even moment $\mu_r(t)$ about its mean value, or $\mu(t) = 0$, is easily determined to be

$$\mu_r(t) = \nu^{r/2}[1 \cdot 3 \cdot 5 \cdot \cdots \cdot (r-1)]/[(\nu - r + 2)\cdots(\nu - 2)] \tag{4-101}$$

where r is even only.

The variance $\sigma^2(t)$ of Student's t is

$$\mathrm{Var}(t) = \sigma^2(t) = \nu/(\nu - 2),\ \nu \geq 2. \tag{4-102}$$

The skewness coefficient α_3 of t is

$$\alpha_3(t) = 0 \tag{4-103}$$

and the coefficient of kurtosis α_4 is given by

$$\alpha_4(t) = 3 + 6/(\nu - 4),\ \nu \geq 4. \tag{4-104}$$

From Eq. 4-104 it is seen that the probability distribution of Student's t approaches the normal distribution very rapidly with increasing ν.

Useful percentage points of Student's t for the practicing analyst are given in Table 4-13, which is reproduced from Ref. 5. Reference to the bottom few rows of Table 4-13 indicates just how rapidly Student's t approaches the normal distribution. This observation leads us to record a very useful alteration of Student's t statistic, due to Smith (Ref. 42); this alteration is of much interest and well to remember.

Smith (Ref. 42) notes that since the variance of Student's t, i.e., Eq. 4-102, is really $(n-1)/(n-3)$, instead of using $\nu=(n-1)$ df for the denominator of the sample standard deviation s in Eq. 4-99, one may divide the SS about the sample mean by $(n-3)$ and refer the new or altered t, which we will call t^*, to a table of the standardized normal distribution. Thus instead of calculating t from Eq. 4-99, we calculate the quantity t^* or

$$t^* = (\bar{x}-\mu)\sqrt{n}/[\Sigma(x_i-\bar{x})^2/(n-3)]^{1/2} = t\,[(n-3)/(n-1)]^{1/2} \tag{4-105}$$

and use the tables of percentiles of the unit normal distribution, i.e., only the bottom line of Table 4-13. The accuracy of this approximation for the upper 5% level of significance has been determined by Scott and Smith (Ref. 43) and indicated in Table 4-14.

One notes from the last column of Table 4-14 that for the widely used upper 5% level of Student's t, one may safely use t^*, the practical consequences of which for five or more df are nil indeed! In summary, for the 5% level of Student's t, one may use t^* with the critical value of 1.96 and abandon Student's t table of percentiles.

In this chapter, we discuss the case of continuous variables. The case of discrete variables and the use of count data, especially to compare binomial population parameters, are discussed in Chapter 5.

4-7.2 CONFIDENCE BOUNDS ON THE UNKNOWN NORMAL POPULATION MEAN

Student's t statistics of either Eq. 4-99 or Eq. 4-105 contain only the single nuisance population parameter or mean μ, being free of the unknown σ. Hence for a single random sample of size n drawn from a hypothesized normal population, one cannot only test the assumption that μ takes on a given or stated value, but he can also calculate confidence bounds on the unknown value of the population mean μ. If we test the null hypothesis H_0 that $\mu=\mu_0$, the sample mean and standard deviation along with the assumed value μ_0 of μ are substituted into Eq. 4-99 or Eq. 4-105 to determine whether the observed value of t is significant or not, thereby making a statistically valid judgment on the size of μ_0.

On the other hand, the probability statement

$$Pr[-t_\alpha \le t \le t_\alpha] = 1-2\alpha \tag{4-106}$$

where

t_α = upper α probability level of Student's t

may be inverted to obtain from Eq. 4-99, for example, that the $(1-2\alpha)$ confidence bound on μ is available from the statement

$$Pr[\bar{x}-t_\alpha s/\sqrt{n} \le \mu \le \bar{x}+t_\alpha s/\sqrt{n}\,] = 1-2\alpha. \tag{4-107}$$

Hence for a single random sample drawn from a normal population $N(\mu,\sigma^2)$, we may obtain confidence bounds on both parameters—i.e., confidence bounds or σ^2, or σ, from Eq. 4-64, or I_{ML} from Table 4-6, or I_{SU} from Table 4-7, and the bounds on μ from Eq. 4-107.

In Example 4-4 we determined confidence bounds on the unknown σ for the data of Example 4-1. In Example 4-7 we illustrate the use of Eq. 4-107 to obtain bounds on μ.

Example 4-7:

Use the data of Example 4-1 to calculate 95% confidence bounds on μ.

We have $n=11$, $\bar{x}=1496.36$ ft/s, $s=10.25$ ft/s, and from Table 4-13 the upper $t_{0.025}=2.228$.

Hence by employing Eq. 4-105

TABLE 4-13

PERCENTILES OF THE t DISTRIBUTION (Ref. 5)



ν = df	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576

From *Introduction to Statistical Analysis* by W. J. Dixon and F. J. Massey. Copyright © 1957 by McGraw-Hill Book Company. Used by permission of McGraw-Hill Book Company.

TABLE 4-14

SCOTT AND SMITH'S t APPROXIMATION—(95% LEVEL) (Ref. 43)

ν = df	$t_{0.95}$	$t^*_{0.95}$	$z_{0.95}$	p^* (probability)
5	2.571	1.992	1.96	0.053
10	2.228	1.993	1.96	0.053
20	2.086	1.979	1.96	0.052
60	2.000	1.966	1.96	0.051
∞	1.960	1.960	1.96	0.050

$t^*_{0.95}$ = 95% probability level of t^*
$z_{0.95}$ = upper 5% point of standard normal deviate
p^* = probability level achieved by using t^* as a normal deviate

$$Pr[1496.36 - 2.228(10.25)/\sqrt{11} \leq \mu \leq 1496.36 + 2.228\,(10.25)/\sqrt{11}] = 0.95$$

or

$$Pr[1489.47 \leq \mu \leq 1503.25] = 0.95.$$

We now turn to Student's t test for two samples, which is used for testing the hypothesis that the two samples come from normal populations with the same (equal) mean(s).

4-7.3 STUDENT'S t TEST FOR TWO NORMAL SAMPLES

We see from par. 4-7.1 and especially from Eq. 4-99 that Student's t statistic involves the difference of the sample mean $\bar{x}$ and the unknown population mean μ in the numerator, whereas the denominator is an estimate of the standard deviation of this difference or, simply, of $\bar{x}$. Thus and quite generally, we may extend this principle to the comparison of two samples. In fact, for two samples assumed to be drawn from the same or perhaps two different normal populations, we may establish a Student's t ratio by taking the difference between the two sample means, subtracting from that the difference between the two normal population means, and then dividing by the proper estimate of the standard deviation of the numerator. However, we will encounter several problems of interest in this connection.

Student's t statistic is primarily, at least as covered here, a test concerning equality of population means. The Snedecor-Fisher F statistic was used to test the hypothesis that two population variances are equal. Thus the F test may establish that, based on the ratio of two sample variances, the two population variances are not equal. This would lead to some problems. It can be seen that if the F test justifies the assumption of equality of variances, we may as well pool the two sample variances and obtain a more stable estimate of the standard error of the difference in means. The problem then is how to pool sample variances. Moreover, this is especially the case if the F test negates the equality of population variances. We will make these considerations clearer and more precise with the following definitions of symbols and subsequent treatment.

Let

μ_1 = population mean of first normal population
μ_2 = population mean of second normal population
σ_1 = population standard deviation of first normal population
σ_2 = population standard deviation of second normal population
n_1 = sample size of "first" sample (drawn from first population)
n_2 = sample size of "second" sample (drawn from second population)

$\bar{x}_1$ = sample mean of first sample

$\bar{x}_2$ = sample mean of second sample

$S_1^2 = \sum_{i=1}^{n_1}(x_{i1} - \bar{x}_1)^2$ = SS about the first sample mean

$S_2^2 = \sum_{i=1}^{n_2}(x_{i2} - \bar{x}_2)^2$ = SS about the second sample mean

$s_1^2 = S_1^2/(n_1 - 1)$ = sample variance of first sample based on $(n_1 - 1)$ df

$s_2^2 = S_2^2/(n_2 - 1)$ = sample variance of second sample based on $(n_2 - 1)$ df.

With these symbolic definitions we will proceed in steps to test various hypotheses—especially the two major ones concerning whether $\mu_1 = \mu_2$—first by accepting the hypothesis that $\sigma_1 = \sigma_2$, and then by proceeding to discuss the so-called Behrens-Fisher problem for which it is known or judged that $\sigma_1 \neq \sigma_2$.

4-7.3.1 Student's t for the Case $\sigma_1 = \sigma_2$

Suppose we have two normal samples and either know or have established by the Snedecor F test that $\sigma_1 = \sigma_2 = \sigma$. In this case, we have only to test whether $\mu_1 = \mu_2$ to establish that the two samples come from the same normal population, for then both sigmas would be equal. Student's t test for equality of population means would then be rather straightforward. In fact, we should, based on the F test establishing that $\sigma_1 = \sigma_2$, simply add the two sums of squares S_1^2 and S_2^2 and divide by the total number of df, i.e., $(n_1 - 1)$ plus $(n_2 - 1)$ to obtain the best estimate of the common population variance σ^2. Thus the estimate $\hat{\sigma}^2$ of σ^2 would be the best available quantity

$$\hat{\sigma}^2 = (S_1^2 + S_2^2)/(n_1 + n_2 - 2). \tag{4-108}$$

If we remember that this is the estimate of the variance of an individual observation and that the variance of $\bar{x}_1$ would be $\hat{\sigma}^2/n_1$ and that of $\bar{x}_2$ would be $\hat{\sigma}^2/n_2$, the appropriate Student's t test to judge whether $\mu_1 = \mu_2$ would be

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\hat{\sigma}(1/n_1 + 1/n_2)^{1/2}} \tag{4-109}$$

where we would put $\mu_1 = \mu_2$ or really use only the sample statistic

$$t = (\bar{x}_1 - \bar{x}_2)/[\hat{\sigma}^2(1/n_1 + 1/n_2)]^{1/2}. \tag{4-110}$$

Note that the denominator of Eq. 4-109 is actually the estimated standard error of the numerator $(\bar{x} - \mu_1) - (\bar{x}_2 - \mu_2)$. In this particular case, the two-sample Student's t test, being "robust" or rather insensitive to moderate departures from normality, is really quite powerful in judging whether in fact one may conclude that $\mu_1 = \mu_2$. If we actually judge that $\mu_1 = \mu_2$, we further conclude that the two samples come from the same normal population, or process, and hence there is no superiority of one over the other.

When the F test rejects that $\sigma_1 = \sigma_2$, however, the problem to decide whether $\mu_1 = \mu_2$ even though $\sigma_1 \neq \sigma_2$ becomes much more difficult. We discuss this next.

First, however, let us say a word about calculation of Student's t for the unequal sample size case to avoid the accumulation of rounding error. Since we deal with sums and SS of the sample observations, Eq. 4-108 becomes by expansion

$$t = \frac{(n_1 + n_2 - 2)^{1/2}(n_2\Sigma x_{i1} - n_1\Sigma x_{i2})}{\langle (n_1 + n_2)\{n_2[n_1\Sigma x_{i1}^2 - (\Sigma x_{i1})^2] + n_1[n_2\Sigma x_{i2}^2 - (\Sigma x_{i2})^2]\}\rangle^{1/2}} \tag{4-111}$$

$$= \frac{(n_1 + n_2 - 2)^{1/2}(n_2\Sigma x_{i1} - n_1\Sigma x_{i2})}{[(n_1 + n_2)(n_2 A_{x_1x_1} + n_1 A_{x_2x_2})]^{1/2}}$$

where

$$A_{x_ix_i} = n_i\Sigma x_i^2 - (\Sigma x_i)^2. \tag{4-112}$$

Actually, all of the quantities in Eq. 4-111 may be calculated and stored on many scientific-type pocket calculators; accordingly, Eq. 4-111 is very convenient and accurate for computation of Student's t.

4-7.3.2 The Behrens-Fisher Problem ($\sigma_1 \neq \sigma_2$)

When it is known or otherwise established from the F test that we cannot consider that $\sigma_1 = \sigma_2$—and we still desire to test the hypothesis that $\mu_1 = \mu_2$, or equality of populations—Student's t is not so straightforward. Let us examine this problem now and even for the general case $n_1 \neq n_2$.

There is a very extensive body of literature on the exact solution of the Behrens-Fisher problem; however, we will only suggest some suitable approximate solutions for Army analysts.

Note first that the numerator of t for the Behrens-Fisher problem will be the difference of $\bar{x}_1$ and $\bar{x}_2$, the two sample means. Now the variance of $(\bar{x}_1 - \bar{x}_2)$ is clearly

$$\sigma^2(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \tag{4-113}$$

which certainly may be estimated from

$$\hat{\sigma}^2(\bar{x}_1 - \bar{x}_2) = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}. \tag{4-114}$$

We note that we are not pooling sums of squares but are using them separately to estimate σ_1^2 and σ_2^2 since we judge that $\sigma_1^2 \neq \sigma_2^2$. So far so good, but if we were to take

$$t = \frac{\bar{x}_1 - \bar{x}_2}{[(s_1^2/n_1) + (s_2^2/n_2)]^{1/2}} \tag{4-115}$$

what is the appropriate number of df to enter Student's t tables? Some writers have suggested that we take the number ν df to be

$$\nu \approx \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{s_1^4/[n_1^2(n_1 - 1)] + s_2^4/[n_2^2(n_2 - 1)]} \tag{4-116}$$

as a good approximation.

Alternatively, we note that the quantity

$$s_1^2/n_1 + s_2^2/n_2$$

is a quadratic form in normal variables x_{i1} and x_{i2}, and if we use the approximate chi-square technique of par. 4-4.5, the interested and curious reader may verify by using Eqs. 4-69 through 4-71 that the approximate number of degrees of freedom is

$$\nu \approx \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{s_1^4/[n_1^2\,(n_1+1)] + s_2^4/[n_2^2(n_2+1)]}. \tag{4-117}$$

We will comment on both of these approximations in the sequel, especially by an example, but this will be after we present other techniques.

A somewhat different approach for the unequal variance problem, which develops a maximum value of t in order to determine whether even that value would be significant and others not, is due to Kulkarni (Ref. 44). Kulkarni (Ref. 44) notes that when $\sigma_1 \neq \sigma_2$, the correct value of t in which one is actually interested involves the nuisance parameters σ_1 and σ_2, but it can be put in the form of additive chi-squares as

$$t = (\bar{x}_1 - \bar{x}_2)\,(\sigma_1^2/n_1 + \sigma_2^2/n_2)^{-1/2}\left(\frac{n_1 s_1^2/\sigma_1^2 + n_2 s_2^2/\sigma_2^2}{n_1 + n_2 - 2}\right)^{-1/2}. \tag{4-118}$$

Kulkarni then puts $\sigma_1^2/\sigma_2^2 = y$ in Eq. 4-118 and obtains Student's t as a function of the "variable" y. Eq. 4-118 is then differentiated for y, equated to zero, and the value of y giving the maximum value of t is found. The maximum value of t is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{(s_1 + s_2)/(n_1 + n_2 - 2)^{1/2}}. \tag{4-119}$$

Kulkarni (Ref. 44) then points out that if the maximum value of t in Eq. 4-119 is not significant at the level of the particular percentage point chosen, one can say without regard to the relative sizes of the unknown σ_1 and σ_2 that the null hypothesis $\mu_1 = \mu_2$ tested turns out to be very reasonable indeed.

As a comment, we note that the expected standard error of the difference in averages $(\bar{x}_1 - \bar{x}_2)$ is simply

$$\sigma_{\bar{x}_1 - \bar{x}_2} = (\sigma_1^2/n_1 + \sigma_2^2/n_2)^{1/2} \tag{4-120}$$

whereas for equal sigmas and equal sample sizes $n_1 = n_2 = n$, Eq. 4-120 becomes

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{2}\sigma/\sqrt{n}. \tag{4-121}$$

On the other hand, and for these same assumptions, we can see that the corresponding standard error in the denominator of Eq. 4-119 is

$$\sigma_{\bar{x}_1 - \bar{x}_2} \approx 2\sigma/\sqrt{2n-2} = \sqrt{2}\sigma/\sqrt{n-1} \tag{4-122}$$

which is perhaps surprisingly not much larger for small sample sizes. It would seem, therefore, that the Kulkarni test could be very useful in many practical situations.

We should probably regard Kulkarni's suggestion in Eq. 4-119 as an approximate solution to the Behrens-Fisher problem although it is a good first try, so to speak. Hence and as another approximate solution based on the use of $(n-3)$ as a divisor instead of $(n-1)$ in Eq. 4-105, we will now record the work of Scott and Smith (Ref. 43).

Following the Letter to the Editor of *The American Statistician* by Smith (Ref. 42) and the work of J. B. de V. Weir in Refs. 45 and 46, Nelson (Ref. 47) points out that Weir (Ref. 46) should be credited with the following approximations to the usual two-sample Student's t test for either (1) the case of equal variances or (2) the case of unequal variances. When the F test establishes that $\sigma_1 = \sigma_2$, the two-sample Student's t to use is (Ref. 47)

$$t = t_s = (\bar{x}_1 - \bar{x}_2) \Big/ \left[\frac{S_2^1 + S_2^2}{n_1 + n_2 - 2} (1/n_1 + 1/n_2) \right]^{1/2} \tag{4-123}$$

which presumably could be referred to a table of standard normal percentage points for a sufficiently accurate answer.

On the other hand, when the F test indicates that $\sigma_1 \neq \sigma_2$, the approximate Student's t to use is the quantity (Ref. 47)

$$t = d_s = (\bar{x}_1 - \bar{x}_2) \Big/ \left[\frac{S_1^2}{n_1(n_1 - 3)} + \frac{S_2^2}{n_2(n_2 - 3)} \right]^{1/2} \tag{4-124}$$

(The capital S's, or S_1 and S_2, recall, are the SS about the proper sample means.) See also Adcock (Ref. 48).

Another very useful test for the Behrens-Fisher problem is Cochran's test (CT) covered in Refs. 49 and 50. Cochran's test uses the ratio of the difference between the two sample means and the standard error of this difference as in Eq. 4-115, but it also employs a weighted average or value of the two percentage points of the Student's t based on the two unequal sample sizes if that condition obtains. Thus for the two-sided test that $\mu_1 = \mu_2$, the CT rejects the null hypothesis H_0 that $\mu_i = \mu_2$ if

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{(s_1^2/n_1 + s_2^2/n_2)^{1/2}} > \frac{(s_1^2/n_1)t_1 + (s_2^2/n_2)t_2}{s_1^2/n_1 + s_2^2/n_2} \tag{4-125}$$

where

t_1 = upper $\alpha/2$ percentage point of t for $(n_1 - 1)$ df
t_2 = upper $\alpha/2$ percentage point of t for $(n_2 - 1)$ df.

We see in effect that CT avoids the pooling of variances problem by obtaining a weighted average of two percentage points based on estimated variances of the means $\bar{x}_1$ and $\bar{x}_2$ while also recognizing sample size differences.

Lauer and Han (Ref. 51) have studied rather extensively the power of CT for the Behrens-Fisher problem and find it to be efficient indeed. Also Lauer and Han (Ref. 51) studied especially the use of the preliminary test of significance, or the F test, to judge whether $\sigma_1 = \sigma_2$ and found that CT, after and along with the preliminary test of significance (PTS), provided a good procedure in practice. We believe, therefore, that the Army analyst probably should have some good and extensive use of the procedure using jointly the F test, or PTS, and the CT.

Although we have discussed several test procedures concerning the Behrens-Fisher problem, and hopefully the ones of more immediate interest to the Army analyst,the statistical literature on the subject of exact and approximate solutions is large indeed. Consequently, some readers may desire to develop their knowledge more extensively by using the bibliography and following up on the references included herewith since our account in this chapter has been more or less an introduction to the subject.

At this point, Example 4-8—which makes use of some of the techniques we have discussed for the Behrens-Fisher problem—should be instructive.

Example 4-8:

A standard lot of mechanical time fuzes was reserved for reference purposes. A manufacturer proposed a new fuze and produced 10 prototypes for a comparison test with the old standard lot. In fact, there had been some dissatisfaction with the old reference fuzes. For the comparative test, 12 of the old standard fuzes were assembled to projectiles along with 10 new prototypes, and the 22 rounds were fired alternately from a gun. The results of the firing are given in Table 4-15. From these limited firings is there any evidence that the proposed fuzes are superior to the current standard fuzes? In particular, can it be judged that the new fuzes

have a smaller standard deviation, that the population means of the two fuzes are equal, or better still that the two samples can be considered to have come from the same normal population?

TABLE 4-15

OBSERVED FUZE TIMES

Old Standard, s	New Proposed Fuze, s
5.09	4.85
5.04	4.93
4.95	4.75
4.92	4.77
4.97	4.67
5.15	4.87
4.98	4.67
5.12	4.94
5.23	4.85
4.85	4.75
5.26	
5.16	

Let us refer to the old standard by the designation 1 and that of the new fuzes by the designation 2. Then the pertinent sample sizes, averages, and standard errors are

$$n_1 = 12 \qquad n_2 = 10$$
$$\bar{x}_1 = 5.060 \qquad \bar{x}_2 = 4.805$$
$$s_1 = 0.129 \qquad s_2 = 0.098$$
$$S_1^2 = 0.016641 \qquad S_2^2 = 0.009604.$$

We first note that the sample standard deviation for the current standard fuze of 0.129 s exceeds that of the prototype, which is 0.098 s. Hence we first use the Snedecor-Fisher F test to determine whether the proposed fuze has a smaller population sigma, based on 11 and 9 df, respectively. Here we have

$$F = (0.129)^2/(0.098)^2 = 1.73.$$

Since from Table 4-8 $F_{0.05}(11,9) = 3.10$ approximately, we cannot say that the new or proposed fuze has a smaller sigma although this might be established in larger scale testing.

Now what about the comparison of population means? In this connection—especially since we could not establish the new fuze has a smaller sigma—we might pool the two sample SS about their means to obtain a common estimate of the variance as in Eq. 4-108 and then use Eq. 4-110 as our t test. However, this is indicated in Eq. 4-123, and for illustrative purposes we will proceed as if we had encountered the Behrens-Fisher problem. For a quick test concerning the equality of population means, we may as well use the approximate Student's t of Eq. 4-124. The reader may verify that Eq. 4-124 gives a t value of 4.73, which, when referred to a table of the normal distribution, gives a very highly significant value indeed. Thus we reject the null hypothesis that the means of the populations from which the two samples were drawn are equal and conclude instead that the new or proposed fuzes have a mean lower by $5.06 - 4.81 = 0.25$ s.

Of course, for the existence of a Behrens-Fisher problem instead of the insignificant variance-ratio test we found here, we might, as a point of interest, assume that the standard error of the first sample had turned out to be 0.250 instead of 0.129. Under this assumption the F test would have shown significance, and the new t value based on Eq. 4-124 would have turned out to be 2.93, which is still very highly significant when referred to a table of the normal probability integral. Hence we would still conclude that the first population is higher by 0.25 s.

Any of the order tests of this paragraph could have been used including, for example, the test of Eq. 4-115 with the number of df given by either Eq. 4-116 or Eq. 4-117, or Kulkarni's test of Eq. 4-119 could have been applied as well as the CT of Eq. 4-125. Thus the reader has available several test procedures to examine our conclusions, which were arrived at by using only the approximate Smith-Weir (Refs. 42, 43, 45, and 46) test statistic.

Finally, we record that there is really no problem in using the new or proposed fuzes for reference purposes, since their standard deviation is smaller and calibration could handle the running time mean value problem by correcting for the bias of about 0.25 s.

4-8 INTRODUCTORY DISCUSSION OF DESIGN AND ANALYSIS OF EXPERIMENTS

The common statistical tests of significance—such as Student's t test concerning a hypothesized value of the normal population mean, the Snedecor-Fisher F test, and Student's t statistic for judging whether normal samples establish equality of population means—apply to the cases of either a single sample or only two samples. At least, this is our coverage so far in this chapter, except for the tests of homoscedasticity in par. 4-6. Moreover, this highlight brings us more or less to a point of rather important interest. We see that the significance tests of par. 4-6—including Bartlett's test, Hartley's test, and CT, for example—are general in character since they can really handle the problem of judging homoscedasticity of two or more sample variances although they do not necessarily point out just which population variances are too large or too small. On the other hand, when we consider two-sample tests concerning equality of population means, we come face-to-face with the problem of homoscedasticity again since it simplifies the comparison of means if the equality of variances is established, and the complication of the Behrens-Fisher-type problem does not really arise. Thus we can say that if a comparison of sample variances establishes homoscedasticity, the comparison of population means through an analysis of samples means is more easily conducted. In fact, let us suppose for the moment that we do indeed have the situation of homoscedasticity or, that is, that the several sample variances can be pooled (through the sum of their sums of squares divided by the total number of degrees of freedom), so to speak, to give a common or single value or estimate of population variance. We might then refer to this common variance as the "internal" variance, or the residual variance. This value of residual variance divided by the sample size would give an estimate of the amount the sample means might be expected to vary if some "extraneous" influences did not exist that result in shifting the levels or population means of the categories from which the different samples were originally taken for the experiment. In view of the existence of such a very desirable state of affairs, we could say that the experiment is in "control"—to use quality control terminology—and indeed we have established homogeneity of means or the equality of population means, especially since there is no evidence that the variation of sample means exceeds that expected from chance conditions. However, if homoscedasticity is not established, any proper analysis of the variability among observed sample means becomes more complicated. At any rate, it could be said that the concepts expressed here lead to the statistical field of ANOVA. Although in this particular case we have visualized the analysis of the variation of means as the ANOVA technique, it is nevertheless true that there may be studies about the analysis of variance of variances, or other statistical quantities. Moreover, for the problem of dealing with the analysis of several or many samples, it is easy to see that we have arrived at the point where it could be of extreme importance to know just how and when the samples were taken since the condition may exist where unwanted or unknown variation could have crept into the experiment. Indeed, it is seen in this connection that much thought and effort should have been expended toward orderly planning of the experiment, especially to control unwanted variability, or to design the experiment so that the effect of variability due to extraneous factors could be assessed and stripped out of the experiment through statistical analysis, and the primary variability in which we are interested could be properly studied. Hence there is a need for statistical design and analysis of experiments of all kinds, especially the more complex types of undertakings, because the analysis of variability or variance and the proper design of experimentation go hand-in-hand for best results. Finally, we might well add concerning this broad and important field of statistical endeavor that the number of comparisons or treatments involved will determine the size of the experiment and the arrangement of the experiment to make direct comparisons. Factors contributing to experimental design are the number of trials or sample sizes (depending especially on available data

concerning variability, if they exist) required to possibly bring out superiority of certain treatments, etc.; the equipment to be used in the test, including measuring instruments; the times and dates of the experiment or parts of it; and the layout or grouping of tests; etc.

Since this handbook is dedicated to certain selected topics in experimental statistics for Army analysts and there are many, many good texts or books available on the hundreds of standard experimental designs, along with methods of analysis, we cannot devote the space to any comprehensive coverage of this highly important statistical area. Rather, since statistical designs of experiments and the best analyses to accompany them can easily be found in our references and bibliography at the end of this chapter or in the statistical literature, we must make a severe selection of topics covering the analysis of multiple sample means, especially for the complex problems in the analysis of variance. Thus having already updated some of the problems of estimation, the more common statistical tests of significance, and the like, we must limit this chapter to recommended reading and discussion of a special example.

As stated in introductory par. 4-1, Refs. 1-5 already contain a wealth of useful reference information on the design and statistical analysis of scientific- and engineering-type experiments. Thus a useful background on the planning and analysis of experiments, and special topics associated therewith, is available—especially in Refs. 3 and 4—so there is no point in repeating such basic topics. Moreover, many worked examples are given in Refs. 1-5, and even the subject of transformations to scales where homoscedasticity is assured before the analysis of mean values (or other sample statistics) is also discussed. Hence we recommend that the reader should first use Refs. 1-5 insofar as is possible. Also Ref. 49 by Cochran and Cox is an excellent text and reference book on the design of experiments as are Ref. 52 by Kempthorne, Ref. 53 by Scheffe', and Ref. 54, which contains two volumes by Johnson and Leone.

In addition to a discussion of the nature of experimentation, factorial experiments, randomized blocks, Latin squares, balanced incomplete block designs, and Youden squares, for example, Ref. 3 contains examples illustrating the analysis of some of these designs of experiments. The analysis of a factorial-type experiment on results from a flame test of fire-retardant treatments of fabrics is given in Table 12-5 (p. 12-19) and Table 12-6 (p. 12-20) of Ref. 3. Also many other useful factorial designs of experiments are listed in Ref. 3.

As an example of a randomized block, a two-way classification in the analysis of variance is given for an experiment representing the "conversion gain" of four resistors measured by six test sets for the data listed in Data Sample 12-3.2, p. 13-4 of Ref. 3. "Conversion power" is defined as the ratio of available current-noise power to applied direct current power expressed in decibel units and is a measure of the efficiency with which a resistor converts direct current power to available current-noise power. The analysis of the two-way classification may be used to strip out the variation due to test set measurement (errors) and to assess the variation due to resistors or vice versa. Also resistor efficiency and / or test set level of measurement effects may be assessed.

Ref. 3 lists many balanced incomplete block designs the Army analyst might well use and also many Youden square arrangements. Thus we call attention to the possible usefulness of Refs. 1-5 which, of course, may be supplemented as required by Refs. 49, 52, 53, and 54.

An example of a one-way classification in the ANOVA is given in Table 2-7, and the components of variance are estimated there. This is for an "interlaboratory" type of test showing the importance of a designed experiment for that problem. Another example of a one-way classification in the ANOVA for several observations per cell is given in Example 3-12 and identifies just which testing laboratories should be investigated for their measurements.

Chapter 33, Ref. 55, and Ref. 56 on the original US Army Ballistics Research Laboratories' hand grenade throwing test give rather detailed use of Graeco-Latin Squares in connection with research and development work. Also Chapter 41, Ref. 55, discusses a very unique application of the Latin Square in a combat simulation to study the choice of the best selection of infantry weapons. Owen's handbook (Ref. 57) is a valuable source of statistical information and tables.

With the citation of these few examples, we will devote our attention now to special applications of Army experimental designs. We refer in particular to the use of an experimental design to evaluate subjective-type judgments on proposals submitted for a weapon development competition or to guarantee the best decision concerning competing research and development (R&D) projects, or the like. The examples we will use—prepared by Mr. Paul C. Cox and suggested for inclusion here by Dr. William S. Agee—relates to a statistical procedure for performing an overall analysis of evaluation by board members who rate several proposals in

connection with a procurement process at the White Sands Missile Range. Our particular illustration, however, will apply to the choice of the best among several competing R&D proposals for a weapon system that is to be developed further as required and procured by the Army.

The statistical procedure can be used to show clearly where significant differences occur between different competing industrial companies and also to locate clusters of two or more proposals that possess no real differences. Moreover, a very desirable feature is the capability of the analysis to strip out the variation due to the raters or judges and to get at the problem of assessing differences among the proposals being rated. Such a statistical analysis should provide a convincing justification for the decision maker to negotiate properly with certain of the proposers and not with the borderline ones. This is especially important since it may not be appropriate to predetermine a "passing grade" but instead to determine which proposals fall within competitive ranges as a function of numerical ratings or scores. If there is a significant difference between the top-rated proposal and the next to top one, a very clear selection results, and the first-ranked proposal might be worth the added cost, if any. If, for example, there is no significant difference among the top three proposals, there is no real justification for selecting one of these if one happens to be more costly than the other two. However, if there is a significant difference between the Number 3 proposal and the Number 4 and if Number 4, or a proposal in the same class as Number 4, is less costly than the lowest priced of the top three, a decision must be made upon a trade-off between price and quality. Here we cover only the case in which each member of the evaluation panel places a numerical rating on each and every proposal. For the case where the raters are divided into groups and each group is assigned a portion of the proposals to be rated, a more complicated design of experiment and statistical analysis will have to be conducted.

The case discussed here is a two-way classification in the analysis of variance where n raters are used to study and evaluate k proposals by rating each proposal on a scale of 1 to 100, i.e., to develop scores for the competing proposals. For convenience it is suggested that the proposals be listed in descending order according to their mean scores for the analysis. A good arrangement for the analysis is that of the symbolic matrix of Table 4-16, where the proposals to be rated are designated by P, and raters are designated by R. The scores are represented by A in the body of Table 4-16, and the sums and means of rows and columns, along with the grand sum and grand mean, are given in the margins and the lower right-hand corner. Equations for the sums and means are also listed on Table 4-16.

The suggested form of the actual analysis of variance is given in Table 4-17.

The F ratios for the raters and proposals, along with the proper number of df as indicated, are compared with the corresponding preselected tabular values of the F distribution from Table 4-8, and insignificance or significance of the sources of variation is observed and then judged. If the differences among the raters are significant, it means that some of the raters may give consistently higher or lower grades than some of the other raters. The analysis removes such anomalies from consideration so that a direct comparison is made of the differences among proposals—our primary goal of analysis. If the differences among proposals are significant, excellent grounds exist for judging that there is a real difference between the submitted proposals, and further study of these differences is warranted. In fact, the job then becomes that of placing the proposals in significant groups and of trying to select the superior proposal, if it exists. This problem is addressed in Example 4-9, which covers a numerical analysis. On the other hand, if there is no significant difference among the proposals, as shown by the F ratio for proposals, there is no need for any further analysis because it becomes evident there are no real differences in the merits of the proposals, and it would appear that the award should be based on the matter of price alone.

Finally, a word about the residual, or mean square, error. This residual variance is the unaccounted for variation in our experiment and analysis. This is largely due to variations in the grading of a given proposal by a given rater, which shows perhaps some random variation under repeated scoring, or it could be an interaction effect, i.e., there may be some tendency for grader h to rate proposal j higher (lower) than proposal k, while grader i would rate proposal j lower (higher) than k. Or, there could be other unidentified causes. In some cases it could become desirable to make an analysis of residuals. In any event, the residual variance becomes a rather natural source of unaccounted for variability by which to judge the other contrasts.

Once it has been established on the basis of the ANOVA (Table 4-17) that significant differences exist among the proposals, then further analysis is required to determine just which proposals differ significantly

TABLE 4-16

SYMBOLIC MATRIX—GRADES FROM N RATERS FOR K RESEARCH PROPOSALS

	Proposal						
Rater	P_1	P_2	P_3	...	P_k	SUM	MEAN
R_1	A_{11}	A_{12}	A_{13}	...	A_{1k}	$A_{1\cdot}$	$\bar{A}_{1\cdot}$
R_2	A_{21}	A_{22}	A_{23}	...	A_{3k}	$A_{2\cdot}$	$\bar{A}_{2\cdot}$
...	...	...	...	...	...	...	...
R_n	A_{n1}	A_{n2}	A_{n3}	...	A_{nk}	$A_{n\cdot}$	$\bar{A}_{3\cdot}$
Sum*	$A_{\cdot 1}$	$A_{\cdot 2}$	$A_{\cdot 3}$	...	$A_{\cdot k}$	$A_{\cdot\cdot}$	
Mean**	$\bar{A}_{\cdot 1}$	$\bar{A}_{\cdot 2}$	$\bar{A}_{\cdot 3}$	...	$\bar{A}_{\cdot k}$		$\bar{A}_{\cdot\cdot}$

*Sums: $A_{\cdot j} = \sum_{i=1}^{n} A_{ij}$; $A_{i\cdot} = \sum_{j=1}^{k} A_{ij}$; $A_{\cdot\cdot} = \sum_{j=1}^{k} A_{\cdot j}$

**Means: $\bar{A}_{\cdot j} = \frac{A_{\cdot j}}{n}$; $\bar{A}_{i\cdot} = \frac{\bar{A}_{i\cdot}}{k}$; $\bar{A}_{\cdot\cdot} = \frac{A_{\cdot\cdot}}{nk}$

where

A_{ij} = score or rating by the ith rater on the jth proposal

$A_{\cdot j}$ = sum of ratings by all raters on jth proposal

$A_{i\cdot}$ = sum of ratings given by the ith rater on all proposals

$A_{\cdot\cdot}$ = sum of ratings by all raters on all proposals

$\bar{A}_{\cdot j}$ = mean of ratings by all raters on jth proposal

$\bar{A}_{i\cdot}$ = mean of ratings given by the ith rater on all proposals

$\bar{A}_{\cdot\cdot}$ = mean of ratings by all raters on all proposals

k = number of proposals

n = number of raters

from the others. There are several methods of procedure for this problem, and the one selected here is that of establishing confidence limits about the mean grade μ_{ij} for each of the k proposals, which can be calculated as follows:

$$\bar{A}_{\cdot j} - t_\alpha\sqrt{MSE/n} \leq \mu_{\cdot j} \leq \bar{A}_{\cdot j} + t_\alpha\sqrt{MSE/n}. \tag{4-126}$$

Here, the MSE is divided by the number n of the raters, and t_α is obtained from a table of Student's t, such as Table 4-13, by entering the table with $(n-1)(k-1)$ df and a preselected confidence level. One may graph or otherwise compare the individual confidence intervals against each other. In interpreting graphical plots of confidence limits about the means, one should be careful because the limits may overlap and there may still be a significant difference in mean values.

Recall that we have already established significance between the scores of the proposals as a group, and hence our problem is to divide the original proposals (based on their means) into two or more groups of homogeneous proposals. There are many, many papers and references in the statistical literature concerning this problem; therefore, the entire field cannot be considered here. Rather, we will give only one procedure the Army analyst might use with profit, and that is the Multiple Range Test of Duncan (Ref. 58).* The reader would do well to consult Scheffe' (Ref. 53) also. To determine whether a significant difference exists between

* Note, however, the comments on the Tukey, Scheffe', and Games and Howell multiple comparisons tests in par. 4-11.

TABLE 4-17

ANALYSIS OF VARIANCE OF DATA FROM TABLE 4-16

Sources of Variance	df	*SS*	*MS*	*F*
Raters	$n-1$	$SSR=\frac{\Sigma(A_{i.})^2}{k}-\frac{(A..)^2}{nk}$	$MSR=\frac{SSR}{n-1}$	$FR=\frac{MSR^*}{MSE}$
Proposals	$k-1$	$SSP=\frac{\Sigma(A_{.j})^2}{n}-\frac{(A..)^2}{nk}$	$MSP=\frac{SSP}{k-1}$	$FP=\frac{MSP^*}{MSE}$
Error	$(n-1)(k-1)$	$SSE=SST-SSR-SSP$	$MSE=\frac{SSE}{(n-1)(k-1)}$	
Total	$nk-1$	$SST=\Sigma\Sigma(A_{ij})^2-\frac{(A..)^2}{nk}$		

*An upper tail F test is used to reject any hypothesis that leads to too large a mean square.

SS = sum of squares (about proper mean value)
SSR = sum of squares due to raters
SSP = sum of squares due to proposals
SST = total sum of squares
SSE = sum of squares due to residual or error variance
MS = mean square
MSR = mean square for the raters
MSP = mean square for the different proposals
MSE = mean square for the error or residual variance term ("error of measurement" for the experiment)
F = Snedecor-Fisher F ratio
FR = F ratio of raters to the residual mean square
FP = F ratio of mean square for proposals to mean square error

two specific proposals, i.e., their means, the Duncan Multiple Range Test uses the residual variance or the MSE, the sample size n, and a factor we will call g. This test is based on the quantity

$$g\sqrt{MSE/n} \tag{4-127}$$

which, if exceeded by the difference between two proposal means, whether or not their scores have adjacent ordering, indicates unequal true levels. Thus the Duncan test provides a "gap" test to make a grouping. The quantity g in Duncan's Multiple Range Test is obtained from a table in Ref. 58 for $(n-1)(k-1)$ df and the order of the mean scores to be compared with the test. Hence with Duncan's test, we are able to divide heterogeneous means into homogeneous groups. The methods of Scheffe' (Ref. 53) are considered to be more powerful, however. We now have sufficient statistical procedures to carry out the complete analysis; therefore, we present Example 4-9 as an illustration.

Example 4-9:

The Army has a crash development program to field a new antitank guided missile (ATGM), referred to as the "WOW" ATGM. Detailed proposals have been invited from six reputable contractors, and a special evaluation panel of five experts has been convened to rate the six proposals. The numerical ratings or scores of the individual experts on each of the six proposals are given in Table 4-18. Is there any evidence that one contractor is superior to the others in this competition?

We will answer this question by making an ANOVA of the scores of the five raters or experts. Note in Table 4-18 that the sums and means of column and row scores are given; the proposal means are ranked for convenience, and the differences between adjacent ranked proposal means are listed at the bottom of the table.

The ANOVA of the scores is given in Table 4-19. Note that the F ratios for both the proposals and the raters are very highly significant using the df indicated and for the upper 5% level of significance from Table 4-8. We conclude, therefore, that the variation among raters and that among proposal ratings cannot under any circumstances be attributed to chance occurrences, and hence we need to continue the analysis to try to determine the superior proposal, if any.

As it turns out, the MS for proposals is greater (more than double) than that for the raters. This would seem to be a desirable condition, showing perhaps that the raters are able to perform a good job of evaluating the proposals with acceptable precision. Moreover, the MSE is only 2.04, or the standard error of the unaccounted for variation in the experiment is only about 1.4 points—an acceptable value indeed. The significant difference among the raters demonstrates, as we indicated earlier, that some raters are consistently higher or lower in their ratings, but this certainly seems to be of little importance because the analysis of variance strips these effects out and accomplishes a more direct comparison of differences among the contractor proposals. Thus we see the sensitivity and usefulness of the ANOVA technique.

If there had been no significant difference among the proposal ratings, any observed numerical differences would have been attributed to chance and the contractor would have been selected on the basis of price and not superior technical merit. Since, however, we have observed quite a significant variation among proposal scores, we should proceed to determine homogeneous groupings. For the jth proposal, confidence limits on the true unknown mean can be calculated with the aid of Eq. 4-126. This has been done for both the 95% and the 90% confidence limits for the six proposals, and the results are given on Table 4-20.

One may note from Table 4-20 that for both the 95% and the 90% confidence limits, there is some overlapping of limits for Proposals 1 and 2. There is a considerable amount of overlapping of the confidence limits for Proposals 3, 4, and 5 but hardly any overlapping of limits for Proposals 2 and 3 except a small amount for the 95% limits. Finally, Proposal 6 very definitely appears to be the poorest of all.

A graph showing the confidence limit calculations for Table 4-20 is depicted on Fig. 4-1. The graph shows very clearly that Proposal 6 is in a low class by itself, that perhaps Proposals 1 and 2 should be in the same or top group, and that Proposals 3, 4, and 5 belong in a group of their own. We also see that it is necessary to proceed with the Multiple Range Test of Duncan since Fig. 4-1 shows some overlap.

There is a significant difference between adjacent proposal means if the difference exceeds the quantity given in Eq. 4-127. For 20 df Ref. 58 gives $g = 2.439$ for the 10% level of significance and $g = 2.950$ for the 5% level. This means that the calculations of Eq. 4-127 turn out to be a difference critical value of 1.56 for the 10% level and 1.88 for the 5% level. A check with the mean values for the proposals in Table 4-18 reveals no significant differences between P_1 and P_2, a significant difference between P_2 and P_3, no significant difference between P_3 and P_4 or P_4 and P_5, but a very highly significant difference between P_5 and P_6. Therefore, it would seem that a good procedure would be to negotiate with P_1 and P_2 although it might be desirable to negotiate perhaps with the top five proposers if cost considerations have great weight and technical achievements are satisfactory. There surely seems to be sufficient grounds for dropping Proposal 6 from any further consideration unless a very definite technical relationship between a score of, for example, 70, and acceptability of the system for Army use has been established and cost considerations for Proposal 6 outweigh other matters. It seems clear also that if Proposal 3 is included in the negotiations, Proposals 4 and 5 should be included also. Finally, it is possible that some type of trade-off between technical merit and price could be encountered and that such a relationship also could be established through the use of the ANOVA technique.

TABLE 4-18

SCORES FOR SIX WOW PROPOSALS BY FIVE RATERS

	Proposals							
Rater	P_1	P_2	P_3	P_4	P_5	P_6	$SUM\ A_{i\cdot}$	$MEAN\ A_{i\cdot}$
R_1	84	81	78	78	75	74	470	78.3
R_2	79	80	74	72	73	67	445	74.2
R_3	84	81	79	77	76	70	467	77.8
R_4	75	76	75	74	72	66	438	73.0
R_5	78	75	75	73	72	66	439	73.2
Sum $A_{\cdot j}$	400	393	381	374	368	343	$A_{\cdot\cdot}$ = 2259	
Mean $A_{\cdot j}$	80.0	78.6	76.2	74.8	73.6	68.6	$\bar{A}_{\cdot\cdot}$ = 75.3	
Difference in Means		1.4	2.4	1.4	1.2	5.0		

TABLE 4-19

ANALYSIS OF VARIANCE OF SCORES FOR WOW PROPOSALS

Sources of Var	df	*SS*	*MS*	*F*
Proposals	5	409.1	81.82	40.11
Raters	4	160.5	40.13	19.67
Error	20	40.7	2.04	
Total	29	610.3		

TABLE 4-20

CONFIDENCE LIMITS FOR MEAN SCORES FOR SIX PROPOSALS

		95%		90%	
Proposal	Mean	LCL*	UCL*	LCL	UCL
1	80.0	78.7	81.3	78.9	81.1
2	78.6	77.3	79.9	77.5	79.7
3	76.2	74.9	77.5	75.1	77.3
4	74.8	73.5	76.1	73.7	75.9
5	73.6	72.3	74.9	72.5	74.7
6	68.6	67.3	69.9	67.5	69.7

* LCL = lower confidence limit

* UCL = upper confidence limit

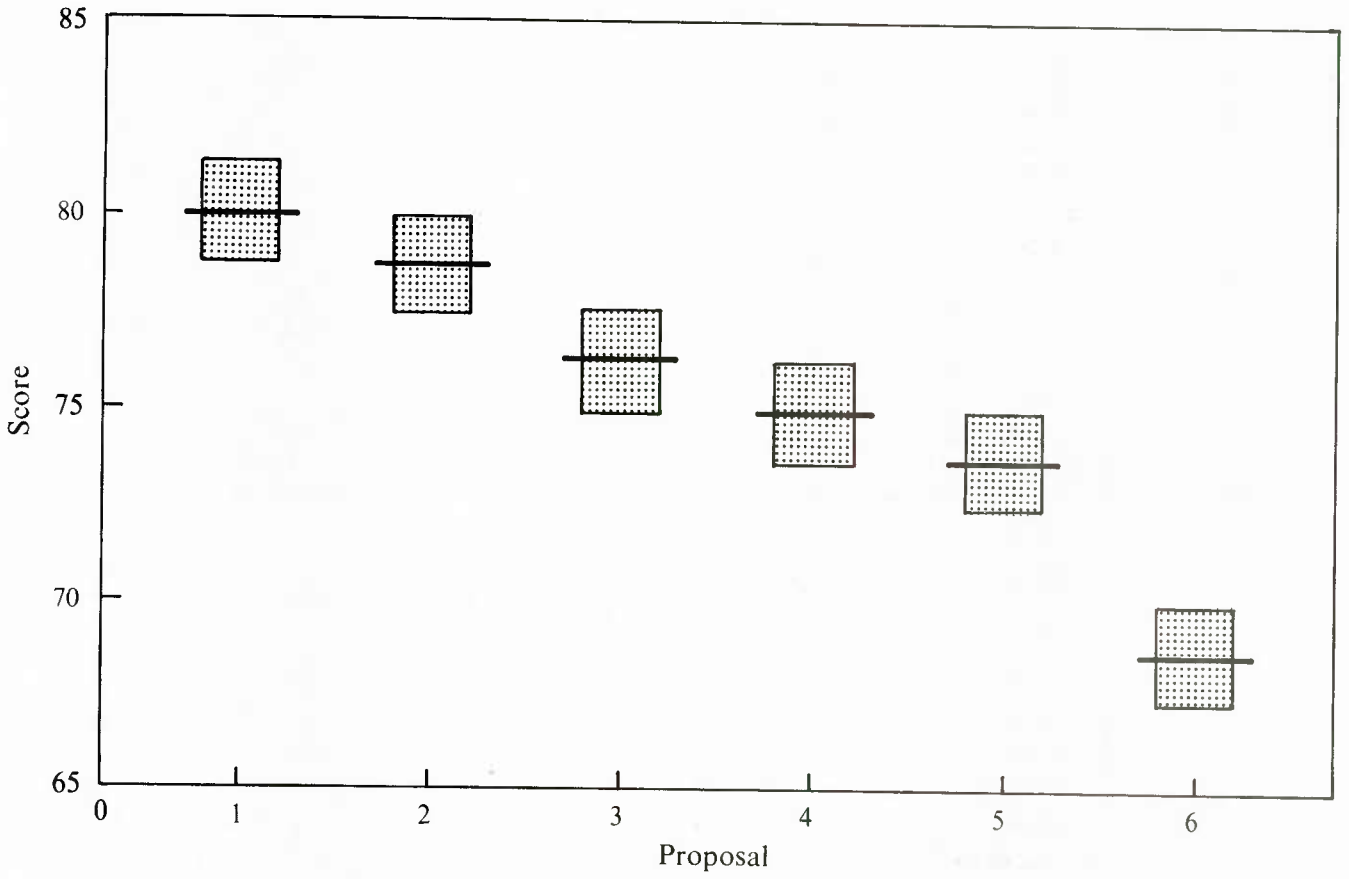


Figure 4-1. 95% Confidence Limits for Each Proposal

We conclude that the design of experiments and the ANOVA technique may have much to offer in contributing to daily decisions in which the Army may be involved because these statistical techniques may be used to quantify the tasks in a superior way. Moreover, the ANOVA technique provides a most efficient way of handling the ever-present and critical problem of wide variation in subjective ratings or judgments.

As a cautionary note to the reader, we record that in our example we have gone ahead with a direct ANOVA without concern about the assumptions of normality or transforming the ratings or count data to another scale or measurement to satisfy normality assumptions.

4-9 COMBINATION OF OBSERVED TAIL AREA PROBABILITIES FOR INDEPENDENT EXPERIMENTS

Experiments are often repeated, or the analyst may be able to get data concerning the significance of several statistical trials or investigations. In view of this, it becomes desirable to know just how the analyst should proceed in order to make the best use of all available data or significance tests that have been conducted. What usually happens is that the statistician calculates the value of a statistic based on the sample observations, such as Student's t, and then the calculated value is referred to a table of the null distribution of the quantity involved. Thus if we let $f(t)$ be the pdf of the statistic in which we are interested, from the appropriate table we, in effect, find the value of the quantity

$$p_i = \int_{-\infty}^{t_i} f(t)dt \tag{4-128}$$

for the ith experiment or significance test, or the complement of it $(1 - p_i)$. We know from statistical theory, however, that the quantity

$$-2\ln p_i = \chi^2(2) \tag{4-129}$$

or that is, the left-hand side is distributed as chi-square with 2 df noting that the upper limit of the integral in Eq. 4-128 is a random variable under the null hypothesis. Thus for a series of the observed probabilities—such as $p_1, p_2, p_3, \ldots, p_n$—then for k significance tests we may sum the $-2\ln p_i$ and treat that sum as chi-square with $2k$ df. By referring this resulting sum to a preselected percentage level of chi-square, we may determine the statistical significance of the whole series of tests.

This result is illustrated by Example 4-10.

Example 4-10:

In two experiments on the delivery "accuracy" of a proposed high-velocity antitank round of ammunition, the first test resulted in an upper tail area probability of 0.07 giving an inconclusive judgment on the round-to-round dispersion at the 5% significance level; a second sample of 10 rounds was fired in the next test. The results from the second test showed significance at the 5% level, and in fact, the observed upper tail area probability turned out to be 0.03. It is possible to combine the two test results and arrive at a definitive judgment?

We have that $p_i = 0.93$ and $p_2 = 0.97$. Hence

$$\begin{aligned} -2\ln p_1 &= +0.14514 \\ -2\ln p_2 &= +\underline{0.06992} \\ 0.20606 &= \chi^2(4). \end{aligned}$$

By referring to Table 4-5 for df = 4, one sees that the 5% level of chi-square is 0.711, and therefore, that the combination of both tests does indeed produce significance at the 5% level or probability—$1 - p \approx 0.005$.

In this particular analysis of final results from two different experiments, one notes that only 4 df are available for the combined test using chi-square, whereas both original sample sizes did, no doubt, have available more than just 2 df each. The reader, therefore, might suspect that the combined test would be rather insensitive. There is, in fact, some loss in efficiency perhaps, as we note in the accuracy firings referenced that the sums of squares from both tests might be pooled to gain df greater than four. Nevertheless, the tests could have been different in type or character, and it may not always be possible to combine sample statistics as desired. Thus the combined chi-square test does indeed have many potential, important uses.

4-10 THE CHOICE OF SIGNIFICANCE LEVELS FOR MULTIPLE TESTS

When the analyst conducts a single test of significance, he decides upon or preselects the level of significance by which he will judge results—this usually amounts to a 5% or a 1% probability level—and then he carries out the calculations for the test and finally compares the value of the observed statistic with the level chosen. As is well-known, however, even this procedure is not straightforward because there is always the question, "Just what level of significance should be chosen?". If, for example, for the outlier detection tests of Chapter 3 one desired to be very careful so that he would not unerringly reject "good" sample values, he might select the 1% or even the 0.5% significance level. Again, however, if the engineer or physicist were looking for sample fatigue test specimens to examine closely on a metallurgical basis, as an example, the 10% or perhaps even the 25% probability level might be selected. Hence we believe that often some very practical guidance, especially concerning the particular physical situation, is of considerable value in the selection of even a probability level for a single statistical test. We therefore urge that the practicing Army statistician come to grips with such a complex problem; by so doing he may well be able to arrive at the best practical solution.

Another important problem for the practicing statistician relates to the choice of significance levels for multiple tests or a series of tests. For example, in the treatment of outlying observations in Chapter 3, we noted the need or temptation to apply outlier tests for a single discordant sample observation, then to test the next suspected sample value after rejection of the first outlier, and so on. Obviously, if we initially chose the 5% level or the 1% level of significance and conducted several significance tests for outliers, the resulting level of probability would change radically from the 5% or 1%, etc., level originally selected. Therefore, one has to exercise care in the choice of percentage points for several tests so that the overall level of statistical significance will be controlled to the desired probability level. This leads us to another complex problem, i.e., the question of the proper choice of a significance level for each test during the course of multiple testing.

Suppose there are m significance tests and the ith test is made at the significance or probability level α_i, say, where we refer to the upper tail area or the pertinent probability distribution. Then the overall significance of the m tests is certainly less than or equal to $\Sigma\alpha_i$. (Similarly, if we were dealing with m confidence intervals, each with confidence $(1-\alpha_i)$, the overall confidence level would be greater than or equal to $(1-\Sigma\alpha_i)$.) Usually, the α_i are taken equal to α/m, where α is the desired (upper) significance level (or the desired confidence level $(1-\alpha)$). A good way of handling problems concerning the probability that one or more of the events will happen, or the "union" U of the sets, is to use the so-called Bonferroni inequalities for the more complex probability calculations, which give either upper and lower bounds or often give exact chances of occurrence. The Bonferroni inequalities are based on a very elementary and basic law for the calculation of the chance of occurrence of at least one of several events, which also would include the occurrence of all of the events simultaneously.

Let there be n events of interest—which are designated by $A_1, A_2, \ldots, A_n$—and let the occurrence of at least one of these events be designated by UA_i. Then, the chance that at least one of the A_i will occur is given by

$$Pr[UA_i] = \Sigma Pr[A_i] - \sum_{i<j}\sum Pr[A_iA_j] + \cdots + (-1)^{n-1} Pr[A_1A_2 \cdots A_n]. \tag{4-130}$$

The right-hand side (RHS) of Eq. 4-130 is very useful because the sum of an odd number of terms of the RHS gives an upper bound on the probability of at least one of the events, and the sum of an even number of terms on the RHS of Eq. 4-130 provides a lower bound of the left-hand side (LHS). Moreover, the sharpness of the bounds increases with the number of terms included. This concept leads to sets of inequalities on lower and upper probabilities of occurrence, which are widely referred to as the Bonferroni inequalities, the first of which is

$$\Sigma Pr[A_i] - \sum_{i<j}\sum Pr[A_iA_j] \leq Pr[UA_i] \leq \Sigma Pr[A_i]. \tag{4-131}$$

One continues the process of placing an even number of terms on the left and an odd number of terms on the right—bracketing the $Pr[UA_i]$—in order to obtain bounds as close as he may desire to the "exact" probability of at least one event. Thus the Bonferroni inequalities are now widely used and, in fact, are necessary in many probability calculations. One additional remark should be made, however—namely, the RHS of Eq. 4-131 can, in many cases, exceed unity. When that happens, the RHS or sum in Eq. 4-131 must be replaced by or limited to unity.

The Bonferroni inequalities, or improvements over it, have been used to study the problem of the choice of significance levels for individual tests in a series of several or multiple experiments. In fact, they often lead to the choice for m tests of a significance level equal to α/m for each individual significance test. We will now restrict our remarks to the use of multiple Student's t tests since they are widely used in practice or applications.

If Student's t test is used, for example, in a one-way ANOVA to make confidence interval statements for m contrasts among, say, k population means, then to assure a significance level of less than or equal to α, the Bonferroni inequalities lead to the use of a significance level for each individual test of $[1-\alpha/(2m)]$ for the two-sided tests. Dunn (Ref. 59) has pointed out, however, that a slightly more powerful test would use $0.5 + 0.5(1-\alpha)^{1/m}$, for each significance level, it being slightly smaller. (These last two quantities are left to right areas.)

Unfortunately, the state of the art has not reached the point that the best procedures for selecting individual significance levels are now available for all the important statistical tests when multiple tests are performed. Rather, it may be necessary to consider each application in appropriate detail. All we can say here is that, generally, for one-sided tests we might use a significance level of α/m for each individual test, whereas for two-sided tests we suggest the use of a significance level of $\alpha/(2m)$ although such a procedure may often be "off the mark". Perhaps and hopefully, these recommendations may not be too poor for current practice until refuted by further research.

Bailey (Ref. 59) gives tables of the Bonferroni t statistic for the 5% and 1% probability levels and for a wide range of df. Some extended tables of t and chi-square for Bonferroni tests with unequal error allocation have been provided by Dayton and Schafer (Ref. 60). These publications and references should be of value to interested readers.

4-11 SOME FURTHER COMMENTS

Although some topics are covered in sufficient detail in this chapter—so that the Army analyst may use certain of the techniques to advantage—we have only touched on the extremely extensive subject of multiple comparison procedures, which are critical in ANOVA tests after having observed a significant Snedecor F ratio for more than two treatment effects, blocks of the experiments, etc. Therefore, in the interest of providing the reader with further references to study and apply as needed to his particular experimental problems, we urge that he review Refs. 61-68 because they should be most helpful. Many of the advances in the problem of multiple comparisons are based more or less on an original unpublished manuscript of Tukey (Ref. 68), which has been widely distributed. Many of the multiple comparison procedures use the Studentized range (Ref. 69) for testing the equality or inequality of population means for k samples in an analysis of variance after observing a significant F ratio. Keselman and Rogan (Ref. 70) recently made an extensive study of comparisons of the modified Tukey and Scheffe' methods of multiple comparisons for pair wise contrasts and recommended the Games and Howell (Ref. 64) modification of the Tukey multiple comparison test for pair wise comparisons of means because the Games and Howell procedure not only controlled the Type I error at or below the nominal size but did so for unequal sample sizes and equal or unequal variances. At the same time it was apparently the more powerful procedure. In view of this, it seems appropriate to record the Games and Howell procedure. It consists of testing the difference between the ith and jth sample means of k such treatments based on the statistic

$$q = (\bar{x}_{i\cdot} - \bar{x}_{j\cdot})/(s_i^2/n_i + s_j^2/n_j)^{1/2} \tag{4-132}$$

where this is simply the difference in sample means of interest divided by individual estimates of their variances (which are summed and the square root taken) and q is the Studentized range even though a t-type ratio is designated as q since a table of the Studentized range (Ref. 69, for example) is entered to test for significance. The parameters to enter the Studentized range table are k for the total number of sample means, and the df are given by

$$\nu = (s_i^2/n_i + s_j^2/n_j)^2/[(s_i^2/n_i)^2/(n_i - 1) + (s_j^2/n_j)^2/(n_j - 1)]. \tag{4-133}$$

(Note in this connection that the number of df is precisely that of Eq. 4-116, which was suggested for the t test in the Behrens-Fisher problem for two sample means.) Again, however, we remind the reader that tables of the Studentized range (Ref. 69) are used for the multiple comparison test. Finally, we suggest that the reader study the references thoroughly to become sufficiently expert in his applications.

Another matter of importance concerning ANOVA procedures relates to the subject of transformations of the original data in an attempt to guarantee homoscedasticity and normality along with the frequent case of unequal sample sizes. For this problem, Refs. 71-79 will be of much use to the practicing statistician; the details of transformations of all kinds and their adequate behavior are covered by the referenced authors. Fuchs' paper (Ref. 79) is recent (1978) and hence should be more or less current on such matters. The selection and proper use of transformations in the ANOVA are also extremely important topics that cannot be treated here.

For a pertinent and interesting discussion of the relation between science and statistics in general, see Box (Ref. 80).

For an excellent presentation and fairly introductory account of experimental design procedures of much value to Army analysts, study Box, Hunter, and Hunter (Ref. 81).

4-12 SUMMARY

We have recorded in this chapter a special selection of statistical topics to update the 1969 *Experimental Statistics* Handbooks (Refs. 1-5). The subjects covered include some noteworthy topics on estimation, especially unbiased estimation of the normal population standard deviation based on the sample standard errors with $(n - 1)$ df or the divisor n and also the sample MD and sample range. The idea of efficiency is discussed as is the concept of MSE of estimates. Some moment properties of use to the statistician are

included since they may be of fundamental use in many applications. The relationships between the chi-square, binomial, and Poisson distributions are recorded; and the chi-square distribution and its many, many important applications are covered to a considerable extent. Several methods of estimating confidence bounds on the population variance are discussed for the chi-square methodology. The approximate chi-square distribution is introduced for possible use by the Army analyst, and the Snedecor-Fisher variance ratio or F distribution is discussed rather extensively. Significance tests for the equality of many population variances are presented for the up-to-date methods, and the comparison of tests of homoscedasticity is also covered in sufficient detail. Student's t distribution for a single sample mean and for two sample means is thoroughly addressed as is the Behrens-Fisher problem for testing equality of means when one is faced with the inequality of variances for the two samples.

The subject of ANOVA in general for several or many means, as well as the design of all types of experiments and current methods of analysis, was not undertaken in this chapter. Rather, we have recorded some comments of interest and have given a technique using the two-way classification in the ANOVA to rate and rank proposals or to make other types of subjective judgments. The advantage of the ANOVA as presented here is the elimination of the variation among raters and thereby assessing the proposal ratings directly. The ranking of proposals, or the division of them into appropriate groups, is also treated.

Combination of observed tail area probabilities from several experiments is treated, and the choice of significance levels for multiple tests is also discussed, including the use of Bonferroni inequalities.

Although multiple comparison procedures and transformations of data to various scales of measurement for applying ANOVA techniques were not included in this chapter, we nevertheless give a sufficient number of references so that the reader may proceed with further study for his particular applications.

Several examples are given to illustrate the applications of the statistical techniques discussed.

REFERENCES

1. AMCP 706-110, Engineering Design Handbook, *Experimental Statistics, Section 1, Basic Concepts and Analysis of Measurement Data*, December 1969.
2. AMCP 706-111, Engineering Design Handbook, *Experimental Statistics, Section 2, Analysis of Enumerative and Classificatory Data*, December 1969.
3. AMCP 706-112, Engineering Design Handbook, *Experimental Statistics, Section 3, Planning and Analysis of Comparative Experiments*, December 1969.
4. AMCP 706-113, Engineering Design Handbook, *Experimental Statistics, Section 4, Special Topics*, December 1969.
5. AMCP 706-114, Engineering Design Handbook, *Experimental Statistics, Section 5, Tables*, December 1969.
6. Frank E. Grubbs, *Statistical Measures of Accuracy for Riflemen and Missile Engineers*, Edwards Brothers, Ann Arbor, MI, 1964.
7. Edward E. Cureton, "Unbiased Estimation of the Standard Deviation", The American Statistician **22**, 22 (February 1968).
8. Ben W. Bolch, "More on Unbiased Estimation of the Standard Deviation", The American Statistician **22**, 27 (June 1968).
9. John Gurland and Ram C. Tripathi, "A Single Approximation for Unbiased Estimation", The American Statistician **25**, 30-2 (October 1971).
10. Dinesh S. Bhoj, "Unbiased Estimation of σ^2", Letter to the Editor of The American Statistician **26**, 59-60 (October 1972).
11. R. A. Fisher, "A Mathematical Examination of the Methods of Determining the Accuracy of an Observation by the Mean Error and by the Mean Square Error", *Monthly Notices*, Royal Astronomical Society of London **80**, 758-70 (1920).
12. AMCP 706-109, Engineering Design Handbook, *Tables of the Cumulative Binomial Probabilities*, June 1972.
13. E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians* I, Cambridge University Press, Cambridge, England, 1954.

14. R. F. Tate and G. W. Klett, "Optimal Confidence Bounds for the Variance of a Normal Distribution", Journal of the American Statistical Association **54**, 674-82 (September 1959).
15. Jerzy Neyman, "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability", *Philosophical Transactions of the Royal Statistical Society, A*, No. 236, 333-80 (1937).
16. Arthur Cohen, "Improved Confidence Intervals for the Variance of a Normal Distribution", Journal of the American Statistical Association **67**, 382-7 (June 1972).
17. Frank E. Grubbs, "Approximate Circular and Noncircular Offset Probabilities of Hitting", Operations Research **12**, 51-62 (January-February 1964).
18. E. B. Wilson and M. M. Hilferty, "The Distribution of Chi-Square", *Proceedings of the National Academy of Sciences*, Washington, DC, 684-8 (1931).
19. DARCOM-P 706-101, Engineering Design Handbook, *Army Weapon Systems Analysis, Part One*, November 1977.
20. Karl Pearson, Ed., *Tables of the Incomplete β-Function*, Cambridge University Press, Cambridge, England, 1934.
21. M. S. Bartlett, "Properties of Sufficiency and Statistical Tests", *Proceedings of the Royal Society, A*, May 1937, 268-82 (May 1937).
22. M. S. Bartlett and D. G. Kendall, "The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation", Supplement to the Journal of the Royal Statistical Society **8**, 128-38 (1946).
23. W. G. Cochran, "The Distribution of the Largest of a Set of Estimated Variances as a Fraction of Their Total", Annals of Eugenics **11**, 47-52 (1941).
24. H. O. Hartley, "The Maximum *F*-Ratio as a Shortcut Test for Heterogeneity of Variance", Biometrika **37**, 308-12 (December 1950).
25. J. H. Cadwell, "Approximating to the Distributions of Measures of Dispersion by a Power of χ^2", Biometrika **40**, 336-46 (December 1953).
26. E. S. Pearson and H. O. Hartley, Eds., *Biometrika Tables for Statisticians*, 3rd ed., Cambridge University Press, Cambridge, England, 1966.
27. R. E. Glaser, "Exact Critical Values for Bartlett's Test for Homogeneity of Variances", Journal of the American Statistical Association **71**, 488-90 (June 1976).
28. W. J. Dixon and F. J. Massey, Jr., *Introduction to Statistical Analysis*, 3rd ed., McGraw-Hill Book Company, Inc., New York, NY, 1969.
29. H. A. David, "Upper 5% and 1% Points of the Maximum *F*-Ratio", Biometrika **39**, 422-4 (December 1952).
30. Peter S. Gartside, "A Study of Methods for Comparing Several Variances", Journal of the American Statistical Association **67**, 342-6 (June 1972).
31. R. J. Beckman and G. L. Tietjen, "Upper 10% and 25% Points of the Maximum *F* Ratio", Biometrika **60**, 213-4 (April 1973).
32. C. Chambers, "Extension of Tables of Percentage Points of the Largest Variance Ratio, S^2_{max}/S^2_0", Biometrika **54**, Parts 1 and 2, 225-7 (1967).
33. Erik Harsaae, "On the Computation and Use of a Table of Percentage Points of Bartlett's *M*", Biometrika **56**, 273-81 (August 1969).
34. Paul N. Somerville, "Optimum Sample Size for Choosing the Population Having the Smallest Variance (Large Sample Size Results)", Journal of the American Statistical Association **70**, 852-8 (December 1975).
35. K. M. Lal Saxema, "Interval Estimation of the Largest Variance of *k* Normal Populations", Journal of the American Statistical Association **66**, 408-10 (June 1971).
36. William C. Guenther, "Calculation of Factors for Tests and Confidence Intervals Concerning the Ratio of Two Normal Variances", The American Statistician **31**, 175-7 (November 1977).
37. S. John, "Tables for Comparing Two Normal Variances or Two Gamma Means", Journal of the American Statistical Association **70**, 344-7 (June 1975).
38. M. Samiuddin and M. Atiqullah, "A Test for Equality of Variances", Biometrika **63**, 206-8 (April 1976).

39. Gary L. Tietjen and R. J. Beckman, "Tables for Use of the Maximum *F*-Ratio in Multiple Comparison Procedures", Journal of the American Statistical Association **67**, 581-3 (September 1972).
40. M. W. J. Layard, "Robust Large-Sample Tests for Homogeneity of Variances", Journal of the American Statistical Association **68**, 195-8 (March 1973).
41. Larry E. Toothaker, Jana L. Hicks, and James M. Price, "Optimum Subsample Sizes for the Bartlett-Kendall Homogeneity of Variance Test", Journal of the American Statistical Association **73**, 53-7 (March 1978).
42. T. M. F. Smith, "Defining Student's *t*", Letter to the Editor of The American Statistician **26**, 43-4 (December 1972).
43. A. Scott and T. M. F. Smith, "Interval Estimates for Linear Combinations of Means", Applied Statistics **20**, 276-85 (1971).
44. R. N. Kulkarni, "A Refinement of the *t*-Test for Comparison of Means With Small Samples", Letter to the Editor of The American Statistician **23**, 49 (December 1969).
45. J. B. de V. Weir, Nature **185**, 558 (February 1960).
46. J. B. de V. Weir, Nature **187**, 438 (July 1960).
47. Lloyd S. Nelson, "Standardizing Student's *t*", Letter to the Editor of The American Statistician **27**, 93 (April 1973).
48. C. J. Adcock, "Comparing Two Normal Means", Letter to the Editor of The American Statistician **27**, 245 (December 1973).
49. William G. Cochran and Gertrude M. Cox, *Experimental Designs*, Second Edition, John Wiley and Sons, Inc., New York, NY, 1957.
50. William G. Cochran, "Approximate Significance Levels of the Behrens-Fisher Test", Biometrics **20**, 191-5 (1964).
51. G. N. Lauer and C. P. Han, "Power of Cochran's Test in the Behrens-Fisher Problem", Technometrics **16**, 545-9 (November 1974).
52. Oscar Kempthorne, *The Design and Analysis of Experiments*, John Wiley and Sons, Inc., New York, NY, 1952.
53. Henry Scheffé, *The Analysis of Variance*, John Wiley and Sons, Inc., New York, NY, 1959.
54. Norman L. Johnson and Fred C. Leone, *Statistics and Experimental Design in Engineering and the Physical Sciences*, **I & II**, John Wiley and Sons, New York, NY, 1964.
55. DARCOM-P 706-102, Engineering Design Handbook, *Army Weapon Systems Analysis, Part Two*, October 1979.
56. Frank E. Grubbs and Ellsworth B. Shank, *Analysis of Throwing Tests for the BRL Hand Grenade Study*, BRL Memorandum Report No. 603, April 1952.
57. D. B. Owen, *Handbook of Statistical Tables*, Addison-Wesley Publishing Co., Reading, MA, 1962, pp. 105-7.
58. Olive Jean Dunn, "On Multiple Tests and Confidence Intervals," Communications in Statistics **3**, 101-3 (1974).
59. B. J. R. Dailey, "Tables of the Bonferroni *t* Statistic", Journal of the American Statistical Association **72**, 469-78 (June 1977).
60. C. Mitchell Dayton and William D. Schafer, "Extended Tables of *t* and Chi-Square for Bonferroni Tests With Unequal Error Allocation", Journal of the American Statistical Association, **68**, 78-83 (March 1973).
61. G. E. P. Box, "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I, Effect of Inequality of Variances in the One-Way Classification", The Annals of Mathematical Statistics **25**, 290-302 (1954).
62. S. G. Carmer and M. R. Swanson, "An Evaluation of Ten Multiple Comparison Procedures by Monte Carlo Methods", Journal of the American Statistical Association **68**, 66-74 (1973).
63. Paul S. Games, "Multiple Comparisons of Means", American Educational Research Journal **8**, 531-65 (1971).

64. Paul A. Games and John F. Howell, "Pairwise Multiple Comparison Procedures With Unequal N's and/or Variances: A Monte Carlo Study", Journal of Educational Statistics **1**, 113-25 (1976).
65. Yosef Hochberg, "A Modification of the T-Method of Multiple Comparisons for a One-Way Layout With Unequal Variances", Journal of the American Statistical Association **71**, 200-3 (1976).
66. Clyde Young Kramer, "Extension of Multiple Range Testing to Group Means With Unequal Numbers of Replications", Biometrics **12**, 307-10 (1956).
67. Emil Spjotvoll and Michael R. Stoline, "An Extension of the T-Method of Multiple Comparison to Include the Cases With Unequal Sample Sizes", Journal of the American Statistical Association **69**, 975-8 (1973).
68. John W. Tukey, "The Problem of Multiple Comparisons", Unpublished Manuscript, Princeton University, Princeton, NJ, 1953.
69. Leon H. Harter, "Tables of Range and Studentized Range", Annals of Mathematical Statistics **31**, 1122-47 (1960).
70. H. J. Keselman and Joanne C. Rogan, "A Comparison of the Modified-Tukey and Scheffé Methods of Multiple Comparisons for Pairwise Contrasts", Journal of the American Statistical Association **73**, 47-52 (March 1978).
71. John W. Tukey, "On the Comparative Anatomy of Transformations", Annals of Mathematical Statistics **28**, 602-32 (1957).
72. D. F. Andrews, "A Note on the Selection of Data Transformations", Biometrika **58**, 249-54 (1971).
73. A. C. Atkinson, "Testing Transformations to Normality", Journal of the Royal Statistical Society, Series B **35**, 473-9 (1973).
74. M. S. Bartlett, "The Use of Transformations", Biometrics **3**, 39-52 (1947).
75. G. E. P. Box, "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I, Effect of Inequality of Variance in the One-Way Classification", Annals of Mathematical Statistics **25**, 290-302 (1954).
76. G. E. P. Box, "Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems", Annals of Mathematical Statistics **25**, 484-98 (1954).
77. G. E. P. Box and D. R. Cox, "An Analysis of Transformations", Journal of the Royal Statistical Society, Series B **26**, 211-52 (1964).
78. J. J. Schlesselman, "Data Transformation in Two-Way Analysis of Variances", Journal of the American Statistical Association **68**, 369-83, (1973).
79. Camil Fuchs, "On Test Sizes in Linear Models for Transformed Variables", Technometrics **20**, 291-9 (August 1978).
80. George E. P. Box, "Science and Statistics", Journal of the American Statistical Association **71**, 791-9 (December 1976).
81. George E. P. Box, William G. Hunter, and J. Stuart Hunter, *Statistics for Experimenters*, John Wiley and Sons, Inc., New York, NY, 1978.

BIBLIOGRAPHY

Min-Te Chao and Ronald E. Glaser, "The Exact Distribution of Bartlett's Test for Homogeneity of Variances With Unequal Sample Sizes", Journal of the American Statistical Association **73**, 422-6 (June 1978).

Israel Einot and K. R. Gabriel, "A Study of the Powers of Several Methods of Multiple Comparisons", Journal of the American Statistical Association **70**, 574-83 (September 1975).

R. C. Elston, "An Analogue to Fieller's Theorem Using Scheffés Solution to the Behrens-Fisher Problem", The American Statistician **13**, 39-47 (April 1974).

Enrico T. Federighi, "Extended Tables of the Percentage Points of Student's *t*-Distribution", Journal of the American Statistical Association **54**, 683-8 (September 1959).

Jean D. Gibbons and John W. Pratt, "*P*-Values: Interpretation and Methodology", The American Statistician **29**, 20-5 (February 1975).

William C. Guenther, "Some Probabilities Available From Desk Calculators and Their Relation to Tables", The American Statistician **31**, 41-8 (February 1977).

Russell T. Hurlbut and Douglas K. Spiegel, "Dependence of *F* Ratios Sharing a Common Denominator Mean Square", The American Statistician **30**, 74-8 (May 1976).

Oscar Kempthorne and T. E. Doerfler, "The Behavior of Some Significance Tests Under Experimental Randomization", Biometrika **56**, 231-48 (August 1969).

J. S. Mehta and R. Srinivasan, "On the Behrens-Fisher Problem", Biometrika **57**, 649-55 (December 1970).

J. Pfanzagl, "On the Behrens-Fisher Problem", Biometrika **61**, 39-47 (April 1974).

J. Singh, "Pooling Mean Squares", Journal of the American Statistical Association **66**, 82-5 (March 1971).

M. L. Tiku, "Power Function of the *F*-Test Under Nonnormal Situations", Journal of the American Statistical Association **66**, 913-6 (December 1971).

CHAPTER 5
INTRODUCTION TO SOME MODERN ANALYSES OF CONTINGENCY TABLES

This chapter describes and highlights the more important and useful statistical techniques that have been developed over the last quarter of a century for the purpose of analyzing contingency tables. Contingency tables represent univariate, bivariate, or multivariate distributions of qualitative data or enumerative type data, which are most often cross-classified because there may be some form of dependence between the classifications. It is of primary interest to determine whether the null hypothesis of independence can be upheld. Some of the more modern statistical techniques that have been used to advantage in recent years to analyze contingency tables, or "cross-classified categorical data", are outlined.

The topics covered include especially some of the key developments in the classical chi-square analysis approach and also the more recent and powerful principles of the information theory approach of Solomon Kullback that employs his minimum discrimination information statistics (MDIS) to analyze contingency tables of any order. The use of loglinear models is also introduced.

Some special coverage of the important problem of comparing binomial populations is given in appropriate detail, and techniques for the determination of confidence limits on the difference of two binomial parameters, their ratio, or the odds ratio, are discussed.

Many illustrative examples are also presented as somewhat of a training aid.

5-0 LIST OF SYMBOLS

A = designates process or category A

A = designates the characteristic A

$\bar{A}$ = designates the characteristic "not A"

a = frequency or number of observations classified according to the cell definition of the first row and first column of a 2×2 contingency table

$\bar{a}$ = mean of a

B = designates process or category B

B = designates the characteristic B

$\bar{B}$ = designates the characteristic "not B"

b = number of sample observations in the cell of the second row and first column of a 2×2 contingency table

c = cell frequency for the first row and second column of a 2×2 table, or number of columns for a table

c_g = continuity correction. Yates' $c_g = 0.5$

d = cell frequency for the second row and second column of a 2×2 table

d = difference or distance between population and sample values or two quantities

E_i = expected number of occurrences for the ith category

$f(p)$ = function of the quantity p, usually a probability density function (pdf)

H_0 = null hypothesis

H_1 = first or initial set of marginals used in an analysis

H_2 = second set of marginals, used for analysis, which is included in set H_1

h = specified small quantity

I = amount of information

$I(p{:}\pi)$ = amount of information based on the MDIS π table

$\hat{I}$ = estimate of the amount of information

$m = a + c$ = sample size for the first row of a 2×2 table

N = total for the 2×2 table $= m + n = r + s$

$n = b + d$ = sample size for the second row of a 2×2 table (n sometimes refers to the table total)

n = sample size

O_i = observed number of occurrences for the ith category

P = designation for a probability

P_1 = specific probability (see Eq. 5-28)

P_2 = specific probability (see Eq. 5-28)

p = true unknown proportion of defectives (or nondefectives, or successes, etc., in a binomial population)

$\hat{p}$ = sample estimate of the unknown population parameter p

$p(AB)$ = probability of both A and B occurring (The same joint chance applies to other letters, of course.)

$p(ij)$ = true but unknown probability of occurrence, or population proportion, for an individual belonging to the cell in the ith row and jth column of the table

$p(i.) = pr(x = i)$ = marginal probability for ith row

$p(.j) = pr(x = j)$ = marginal probability for jth column

p_A = lower confidence level of p

p_B = upper confidence level of p

p_c = "control" or "standard" value of p

p_t = "test" or "treatment" value of p

p_1 = population parameter for the first binomial population

p_2 = population parameter for the second binomial population

$p^*(ij)$ = cell probability for the ith row and jth column based on the MDIS

$p^*(i.)$ = ith row probability based on the MDIS

$p^*(.j)$ = jth column probability based on the MDIS

$R = p_1/p_2 = p_t/p_c$ = ratio of p's

R_L, R_U = lower and upper confidence limits of R, respectively

r = number of defectives observed, or $r = a + b$, or number of rows

s = sum of c and d

s_d = sample standard deviation of the difference d

$x(ij)$ = observed frequency for the cell in the ith row and jth column, for $i = 1, \ldots, r$ and $j = 1, \ldots, c$

$x(i.)$ = sum of the $x(ij)$ across the c columns of the ith row

$x(.j)$ = sum of the $x(ij)$ across the r rows of the jth column

$x(..) = N$, sometimes n, = the sum of all the observations within the contingency table

$x(11)$ = observed number of occurrences a for the cross-classification involving A and B in Table 5-6

$x(21)$ = observed number of occurrences given by b for the cross-classification $\bar{B}$ and A in Table 5-6

$x^*(ij)$ = predicted value for the cell in the ith row and jth column, which is determined in accordance with Kullback's MDIS principle

x_1^* = refers to the expected frequency for a set H_1 of given marginals

x_2^* = refers to the expected frequency for a second set H_2 of given marginals which is included in H_1

z = unit or standard normal deviate

z_1 = normal deviate defined in Eq. 5-20, which keeps the sample sigmas separate for two different binomial p's, p_1 and p_2

z_2 = normal deviate represented by d/s_d (difference divided by the standard deviation of that difference) (see Eq. 5-17), which pools the two samples of data to obtain a single estimate of σ

z_α = αth probability level of the deviate z. Often, only the deviate for the upper α probability level is used.

α = probability level, less than 0.50 and usually 0.05 or 0.01

α^* = maximum of $f(p)$

α^+ = value to which the computer is instructed to iterate

$\Delta = p_1 - p_2 = p_t - p_c$ = difference of p's

Δ_L, Δ_U = lower and upper confidence limits of Δ, respectively

$\pi(ij)$ = probability for the cell in the ith row and jth column based on the π table $= 1/(rc)$ for the uniform distribution

σ_a^2 = Var(a) = variance of a

$\hat{\sigma}$ = estimate of sigma, the population parameter

$\chi^2 = \chi^2(\ )$ = chi-square statistic with degrees of freedom (df) indicated within the parentheses

$\psi = p_1(1-p_2)/[p_2(1-p_1)] = p_t(1-p_c)/[p_c(1-p_t)]$ = the odds ratio

ψ_L, ψ_U = lower and upper confidence limits of ψ, respectively

5-1 INTRODUCTION

Chapter 4 dealt with measurements or observations on a continuous scale but not generally with binomial- or count-type data. As we are aware, a very large amount of data from experiments, or many experimental observations, leads to characterization into only two categories. Thus an observation is judged simply as a "success" or a "failure", or "pass" or "fail", "go" or "no go", etc. An example would be firing 10 armor-piercing (AP) projectiles having a striking velocity of, say, 1000 m/s at 9 in. of rolled homogeneous armor plate and observing that two of the projectiles did "defeat", or pass through, the plate. These measurements are considered to be on an attribute scale instead of on a continuous scale as we discussed in Chapter 4. Hence our purpose in this chapter is to discuss methods of statistically analyzing such data in order to arrive at some decision about the unknown population parameters, or for other reasons. In particular, we again will have the problem of analyzing data representing one, or two, or more "samples" of cross-classified data.

In a manner quite analogous to the treatment of continuous-type data, the analysis of attribute data also will involve making inferences from a single sample drawn at random from a binomial population, or we may deal with two or more binomial-type samples and be interested in whether the samples can be considered to have been drawn from the same binomial population. One might think, in this connection, that since binomial populations are described by a single parameter, i.e., the proportion of successes or failures, etc., it would naturally follow that the statistical analysis would be much easier. However, this is not always the case because of the discrete nature of the random variables. In any event, the analysis of enumerative data often may be carried out along somewhat similar lines to that of observations on a con-

tinuous scale, and often the discrete type data are displayed in a layout with cells similar to the analysis of variance (ANOVA) form used for continuous data. In fact, our analysis of proposals in par. 4-8 did indeed involve enumerative data or "ratings", and we carried out an analysis of variance on the original measurements as if we were dealing with continuous-type data. Often, however, such a matrix for discrete or enumerative data would be given in the form of and analyzed as a "contingency" table, but this depends on the categories of interest into which the data fit or are taken or drawn originally. For a contingency table we have a random sample of objects that are cross-classified into two or more attributes, and each attribute may be further divided into two or more categories. Thus there are cells or classifications into which none, one, or more of the observations will logically fit or possess the required attributes. Moreover, for each of the cells or attributes involved, there exists a probability for the whole population under consideration that an individual will belong to that particular cell.* For a sample of observations, we will not know just what the true chance is, and, in fact, we will almost always have to estimate such probabilities or at least have to make some inferences about the true unknown population parameters by testing a hypothesis of interest, which states, for example, that two or more samples come from the same population—i.e., the samples are "equivalent" until such a hypothesis is rejected.

A contingency table represents a sample from a multivalued population, and the two-way table, for example, is simply a matrix of observed frequencies cross-classified with the two characterizations, and the display is by rows and columns of the matrix.** For example, we might represent the number of penetrations and nonpenetrations of armor plate by using two columns and two types of heat treatments of the projectiles by two rows; this establishes a "two-by-two" contingency table. Of course, this idea extends to any number of classifications by rows and columns. In this particular example the statistician may proceed to try to establish whether one heat treatment really makes any difference or is superior to the other heat treatment, etc. The basic treatment and analysis of contingency tables are to be found in almost any standard textbook on statistics. Therefore, our purpose is to discuss some topics on contingency tables of interest to Army analysts. Indeed, we should aim to update the very good account of the analysis of enumerative and classificatory data in Ref. 1. The reader is urged to review first this reference as a basis for proceeding with the contents of the present chapter.

In our coverage we will begin with the concept of a single sample from a single binomial population, proceed to a discussion of two-by-two contingency tables, which are of considerable importance in Army statistical investigations, and finally go on to some coverage of the more complex types of contingency tables.

Before proceeding, however, we should warn the reader that efforts toward any unique or straightforward analysis of even the two-by-two contingency table can be very confusing unless one stops to place several different types of problems in proper perspective before the actual analysis is conducted. Indeed, it becomes very important to know just how samples were drawn or selected, and what they really represent—especially whether row or column totals are "fixed", or whether both row and column totals are fixed, etc.—because this would represent very different conditions or areas of analysis. As we shall learn, it was not until about 1947 that the different types of problems in the analysis of contingency tables were made unmistakably clear.

We now summarize very briefly in par. 5-2 some results related to the drawing of a single random sample of n items from a single binomial population, especially that covered in Refs. 1 and 2.

5-2 SAMPLING A SINGLE BINOMIAL POPULATION WITH A SAMPLE OF SIZE n

Following the notation of Ref. 1, we consider the random drawing of a sample of size n from a binomial population with parameter p representing the true unknown proportion of defectives (or failures) or successes, etc. In Army analyses we will more often deal with "failures" or "defectives" because they are most usually the main focus of interest. However, if we are concerned with high reliability or safety,

*The term "cell" is used here to denote the category or classification into which a response fits.

**The 2×2 contingency table is often referred to as a "double dichotomy", especially for the case where the row and column totals are random numbers.

our concentration might shift somewhat. As a result of drawing the single sample of n, we will find r defectives, or failures, and then our main interest will center around estimating the proportion p of failures in the universe and also around placing confidence bounds on this unknown parameter p. It is well-known in this connection that the maximum likelihood (ML), unbiased estimate of the binomial population parameter is given by $\hat{p}$, where

$$\begin{aligned} \hat{p} &= r/n \\ r &= \text{number of defectives (failures) observed} \\ n &= \text{sample size.} \end{aligned} \tag{5-1}$$

Ref. 1 discusses in some detail the problem of placing confidence intervals on the paramater p, which gives the normal approximation for the sample size n greater than 30. Also Table A-22, Ref. 2, gives some very valuable tables for confidence limits on the proportion p for sample sizes of $n \leq 30$. Table A-24, Ref. 2, which actually is figures, gives curves for the upper and lower confidence limits on p for sample sizes of $n = 50$, 100, 250, and 1000. As a matter of record, the $(1 - 2\alpha)$ confidence limits given in Ref. 1 for n greater than 30 are listed as

$$(r/n) - z_\alpha \sqrt{(r/n)\,(1 - r/n)/n} < p < (r/n) + z_\alpha \sqrt{(r/n)\,(1 - r/n)/n}^{*}. \tag{5-2}$$

A more up-to-date treatment of confidence intervals—one devoted especially to reliability, along with use of the Snedecor-Fisher F distribution, the incomplete beta function ratio, and some other procedures—may be found in Chapter 21, *Army Weapon Systems Analysis, Part I*, Handbook (Ref. 3). Some very useful charts for reading off the upper and lower 95% and 99% confidence limits about the binomial p are given in Ref. 4, and we include these in Figs. 5-1(A) and 5-1(B). Ref. 4 also includes tables of confidence limits for the expectation of a Poisson variable with confidence coefficients of 90%, 95%, 98%, 99%, and 99.8%. The Poisson approximation to the binomial becomes valid for "small" p (or p less than about 0.10), and in applications one usually counts the number of defectives or occurrences, which is small, often without knowing the sample size. The *Biometrika* table of confidence limits for the Poisson parameter (Ref. 4) is reproduced here as Table 5-1.

Often it is desired to estimate the unknown binomial parameter p within a distance or difference of d between the population and sample values. If some prior information of p is available or its size is known approximately, then the sample size equation is given by

$$n = z_\alpha^2\, p\,(1 - p)/d^2 \tag{5-3}$$

where

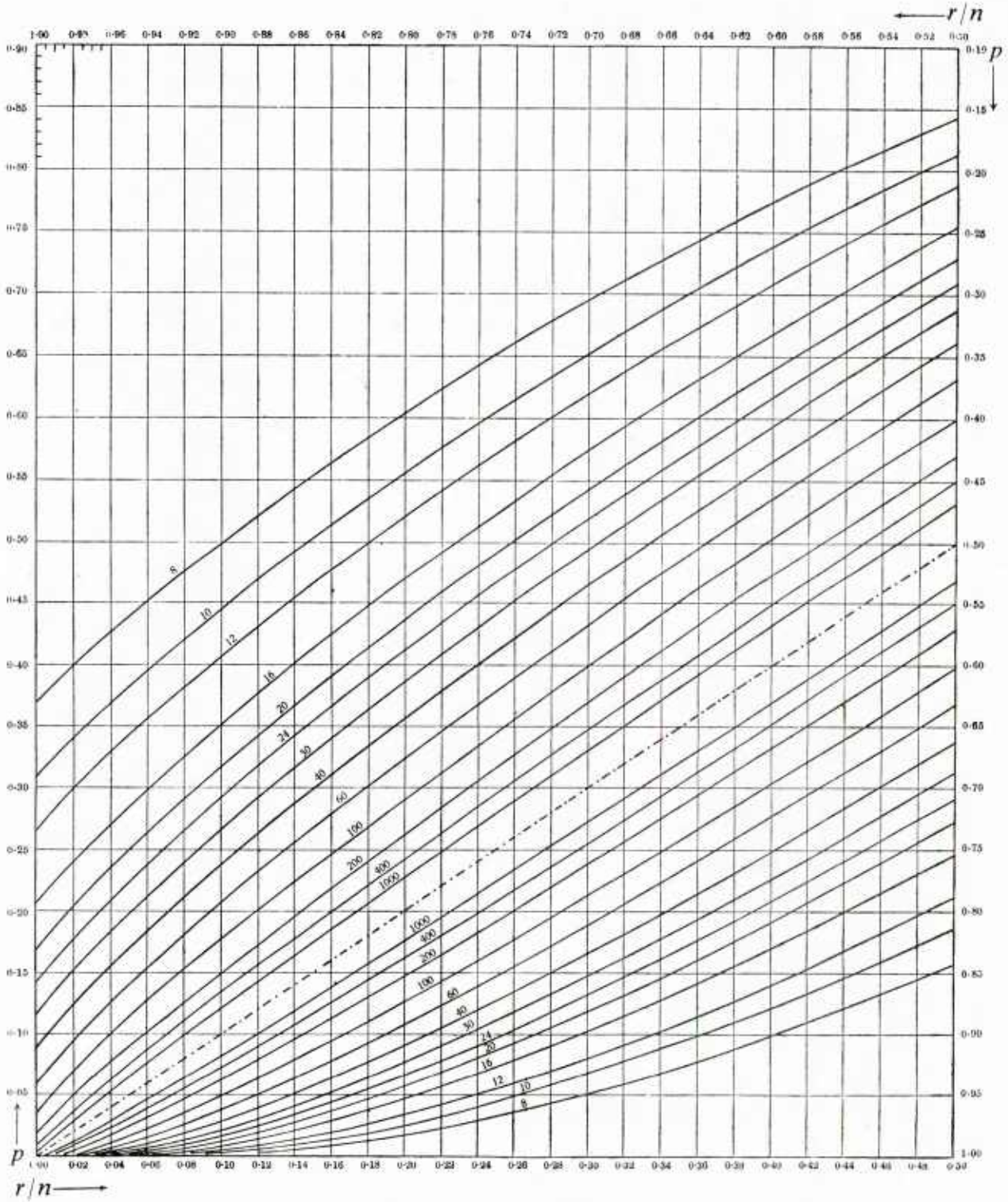
d = difference or distance between population and sample values.

Hence by so determining n, we can say that the sample size n is such that the probability is no more than α that our estimate of p is in error by more than d. In case p is near the value 1/2, Eq. 5-3 reduces to the approximation

$$n \approx z_\alpha^2/(4d^2). \tag{5-4}$$

With this very brief background on sampling a single binomial population, we turn to the comparison of two samples of count type data and especially to the general 2×2 contingency table.

*z_α is used here to denote the upper (positive) α probability level of the standard normal distribution. Some readers will prefer the more precise label $z_{1-\alpha}$. Perhaps it is unfortunate that the two symbols have been used interchangeably in the literature when one is the negative of the other.



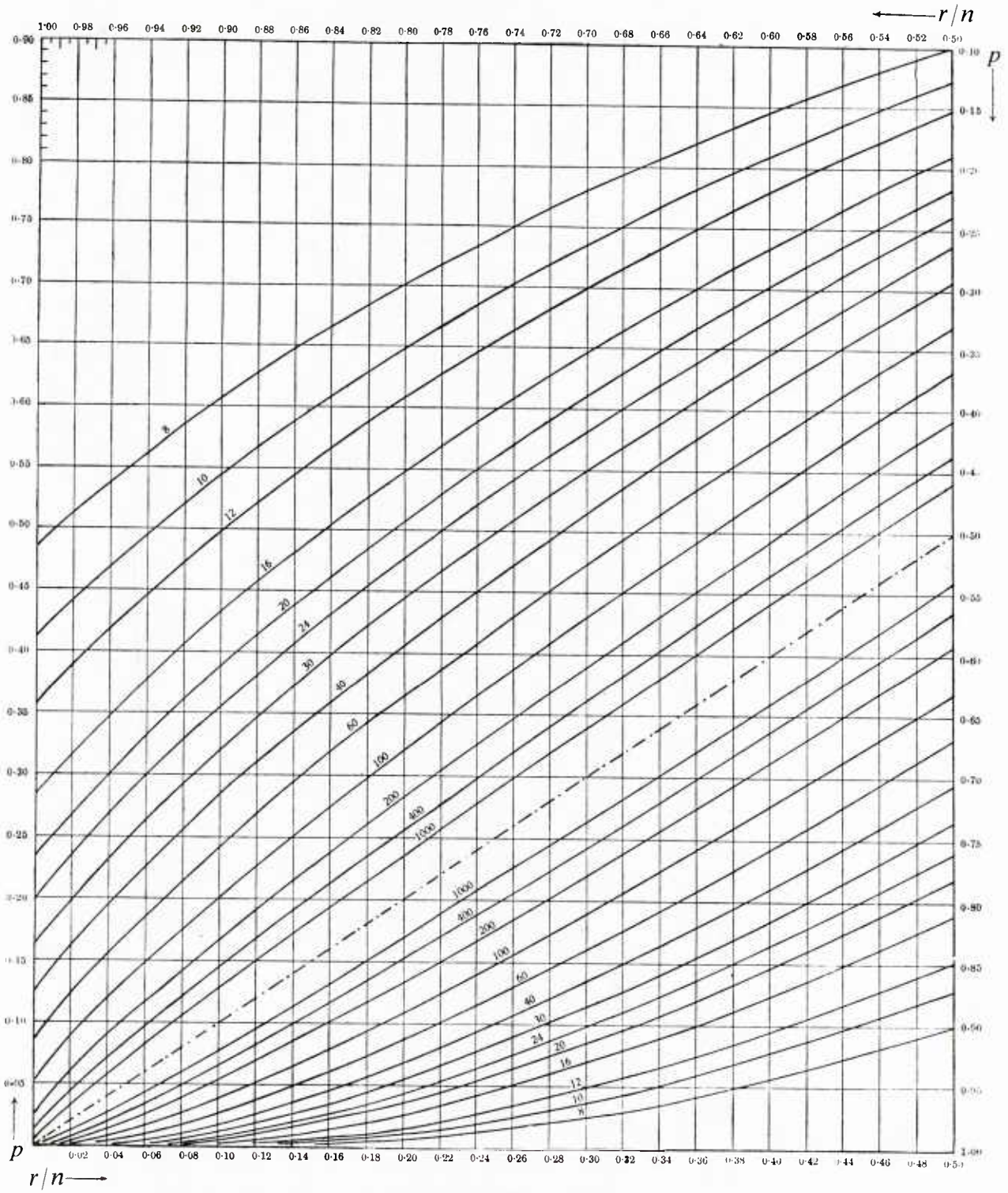
The numbers printed along the curves indicate the sample size n. If for a given value of the abscissa r/n, p_A, and p_B are the ordinates read from (or interpolated between) the appropriate lower and upper curves, then

$$Pr[p_A \leq p \leq p_B] \leq 1 - 2\alpha.$$

Note: The process of reading from the curves can be simplified with the help of the right-angled corner of a loose sheet of paper or thin card, along the edges of which are marked off the scales shown in the top left-hand corner of the chart.

(A) Confidence Coefficient, $1 - 2\alpha = 0.95$

Figure 5-1. Confidence Limits for p in Binomial Sampling, Given a Sample Fraction r/n (Ref. 4)



The numbers printed along the curves indicate the sample size n.

Note: The process of reading from the curves can be simplified with the help of the right-angled corner of a loose sheet of paper or thin card, along the edges of which are marked off the scales shown in the top left-hand corner of the chart.

(B) Confidence Coefficient, $1 - 2\alpha = 0.99$

Reprinted with permission. Copyright © by Biometrika Trustees.

Fig. 5-1 (cont'd)

TABLE 5-1

CONFIDENCE LIMITS FOR THE EXPECTATION OF A POISSON VARIABLE (Ref. 4)

$1-2\alpha$	0.998		0.99		0.98		0.95		0.90		$1-2\alpha$
α	0.001		0.005		0.01		0.025		0.05		α
r	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	r
0	0.00000	6.91	0.00000	5.30	0.0000	4.61	0.0000	3.69	0.0000	3.00	0
1	0.00100	9.23	0.00501	7.43	0.0101	6.64	0.0253	5.57	0.0513	4.74	1
2	0.0454	11.23	0.103	9.27	0.149	8.41	0.242	7.22	0.355	6.30	2
3	0.191	13.06	0.338	10.98	0.436	10.05	0.619	8.77	0.818	7.75	3
4	0.429	14.79	0.672	12.59	0.823	11.60	1.09	10.24	1.37	9.15	4
5	0.739	16.45	1.08	14.15	1.28	13.11	1.62	11.67	1.97	10.51	5
6	1.11	18.06	1.54	15.66	1.79	14.57	2.20	13.06	2.61	11.84	6
7	1.52	19.63	2.04	17.13	2.33	16.00	2.81	14.42	3.29	13.15	7
8	1.97	21.16	2.57	18.58	2.91	17.40	3.45	15.76	3.98	14.43	8
9	2.45	22.66	3.13	20.00	3.51	18.78	4.12	17.08	4.70	15.71	9
10	2.96	24.13	3.72	21.40	4.13	20.14	4.80	18.39	5.43	16.96	10
11	3.49	25.59	4.32	22.78	4.77	21.49	5.49	19.68	6.17	18.21	11
12	4.04	27.03	4.94	24.14	5.43	22.82	6.20	20.96	6.92	19.44	12
13	4.61	28.45	5.58	25.50	6.10	24.14	6.92	22.23	7.69	20.67	13
14	5.20	29.85	6.23	26.84	6.78	25.45	7.65	23.49	8.46	21.89	14
15	5.79	31.24	6.89	28.16	7.48	26.74	8.40	24.74	9.25	23.10	15
16	6.41	32.62	7.57	29.48	8.18	28.03	9.15	25.98	10.04	24.30	16
17	7.03	33.99	8.25	30.79	8.89	29.31	9.90	27.22	10.83	25.50	17
18	7.66	35.35	8.94	32.09	9.62	30.58	10.67	28.45	11.63	26.69	18
19	8.31	36.70	9.64	33.38	10.35	31.85	11.44	29.67	12.44	27.88	19
20	8.96	38.04	10.35	34.67	11.08	33.10	12.22	30.89	13.25	29.06	20
21	9.62	39.38	11.07	35.95	11.82	34.36	13.00	32.10	14.07	30.24	21
22	10.29	40.70	11.79	37.22	12.57	35.60	13.79	33.31	14.89	31.42	22
23	10.96	42.02	12.52	38.48	13.33	36.84	14.58	34.51	15.72	32.59	23
24	11.65	43.33	13.25	39.74	14.09	38.08	15.38	35.71	16.55	33.75	24
25	12.34	44.64	14.00	41.00	14.85	39.31	16.18	36.90	17.38	34.92	25
26	13.03	45.94	14.74	42.25	15.62	40.53	16.98	38.10	18.22	36.08	26
27	13.73	47.23	15.49	43.50	16.40	41.76	17.79	39.28	19.06	37.23	27
28	14.44	48.52	16.24	44.74	17.17	42.98	18.61	40.47	19.90	38.39	28
29	15.15	49.80	17.00	45.98	17.96	44.19	19.42	41.65	20.75	39.54	29
30	15.87	51.08	17.77	47.21	18.74	45.40	20.24	42.83	21.59	40.69	30
35	19.52	57.42	21.64	53.32	22.72	51.41	24.38	48.68	25.87	46.40	35
40	23.26	63.66	25.59	59.36	26.77	57.35	28.58	54.47	30.20	52.07	40
45	27.08	69.83	29.60	65.34	30.88	63.23	32.82	60.21	34.56	57.69	45
50	39.96	75.94	33.66	71.27	35.03	69.07	37.11	65.92	38.96	63.29	50

If r is the observed frequency or count and m_A, m_B are the lower and upper confidence limits, respectively, for its expectation m then

$$Pr[m_A \leq m \leq m_B] \leq 1 - 2\alpha.$$

Reprinted with permission. Copyright © by Biometrika Trustees.

5-3 THE 2×2 CONTINGENCY TABLE WITH EMPHASIS ON COMPARING TWO BINOMIAL POPULATIONS

The 2×2 contingency table probably represents the most important and usual type of statistical analysis of enumerative data with which the Army analyst will likely be confronted. Therefore, it is necessary to discuss 2×2 tables and some pertinent background in some depth. First, we will depict the 2×2 table and discuss some of the possible arrangements of it, and then we will proceed to indicate clearly the different methods of analysis that will be required. Moreover, we will give a fairly complete account of the particular type of analysis that covers the comparison of two binomial populations since this area appears to be of prime interest in applications for the Army analyst or statistician.

5-3.1 THE GENERAL 2×2 CONTINGENCY TABLE

As a basis for preliminary discussion, the general 2×2 contingency table may be put in the form of Table 5-2.

TABLE 5-2

THE GENERAL 2×2 TABLE

	Number Defectives	Number Nondefectives	Total
Process A	a	c	m
Process B	b	d	n
Total	r	s	N

In Table 5-2 we have used letters without subscripts for convenience, and the relations among them are as follows. For Process A there are m items that have been tested or observed, of which the number a of them are classified as "defectives" (or sometimes "successes") and c of them are branded as being "nondefective"; thus $a + c = m$. In a like manner, for Process B, we have b defectives and d nondefectives in a total of n items, observations, etc. The total number of items considered in the experiment is $N = m + n$, whereas the total number of defectives found is $r = a + b$, and the total number of nondefectives is $s = c + d$. Also we have $r + s = N$. Our prime interest in this experiment is to compare Process A with Process B to determine whether any difference really exists or especially to try to judge whether A is superior or inferior to B.

The difficulty with such a simpleminded test or experiment is that so far some of the more important considerations, or points, have not appeared! For example, just how were the number of items, m and n, for the original observations selected? Do, for example, m and n represent random samples from larger categories or different binomial populations? Is the total of N all that interests one and not a larger universe from which N items were possibly drawn? Are m, n, r, and s all "fixed" so that one is only interested in whether the random division into the observed numbers of defectives and nondefectives—or a, b, c,and d—can be judged to represent independence or equality of Processes A and B instead of a low-chance result? Thus it can be seen that it becomes quite important to know the basic physical reasons the experiment was conducted, especially the drawing of items for test, and just what is expected to be learned from the experiment.

Although until about 1947 many statisticians treated 2×2 tables very much alike and mostly used the same test of significance, Barnard (Ref. 5) and Pearson (Ref. 6) began to clear up much of the confusion surrounding contingency tables and brought very striking differences in experimental procedures and analyses of 2×2 tables into sharp focus. They pointed out that one must be careful to distinguish some three different sampling methods of obtaining 2×2 tables and proceed to analyze such contingency tables accordingly. To begin with, one could be interested only in the totality of N items, which have been divided into m and n items for test to determine whether Processes A and B are "equivalent" or "independent" or produce an equal percentage of defectives. Thus in this case we would expect that very nearly $a/m = b/n = r/N$ except for some very random deviations. Barnard (Ref. 5) calls this the "independence" trials experiment, and Pearson (Ref. 6) refers to this case as his "Problem I". No assumption is made con-

cerning how the N individuals were selected, perhaps from a larger universe, and it might be said that one is observing either the presence or absence of a reaction. The first treatment is applied to m items and the second to n items, so that as a result a/m and b/n show reaction to the stimulus applied. This case is also commonly referred to as the "Fisher Exact Test" of 2×2 comparative trials in the statistical literature. The marginal totals m, n, r, and s may all be regarded as fixed, and, since the proportion r/N is known, can one regard the ratios a/m and b/n as being reasonable, in which case there would be no significant differences between Processes A and B?

A second case, and often the one of most importance to the Army analyst, occurs when the m items from Process A have been drawn at random from a large binomial population and n items for Process B have been taken similarly from a second binomial parent. This situation is labeled by Barnard as the "CSM" test* and by Pearson as his "Problem II", in which one is testing whether the proportion of individuals bearing some character—the percent of defectives—is the same in two different populations. It is, however, the well-known problem of comparing the true p's of two different binomial populations to determine whether p_1 for the first parent is equal to p_2 for the second. In the very limited statistical test of significance, we would make a comparison of whether a/m and b/n are sufficiently equal and would reject the null hypothesis of no difference if our statistical test gives a result well into the tails of the pertinent probability distribution used for final judgment.

Finally, there is the third or more general type of 2×2 table in which the random process involves taking N items or individuals from a large population, and each of the N individuals must fall into one or the other of the four cells, or categories, a, b, c, and d. A repetition of drawing another total sample of N individuals would lead to a different set of random numbers falling into cells a, b, c, and d, and also the marginal totals m, n, r, and s would change as well! Pearson (Ref. 6) calls this case his "Problem III", and it actually results in the multinomial distribution as one might easily surmise.**

Although some readers may have difficulty understanding the sharp distinctions Barnard (Ref. 5) and Pearson (Ref. 6) attempt to make, they also may become a bit more confused when it is known that for rather large samples a normal approximation may give sufficiently good results in all three cases! Nevertheless, this is a useful development indeed because the more exact computations become so unwieldy that suitably accurate approximations are welcome and most often must be made. Hopefully, we will be able to clear up some of the difficulties by means of selected examples.

Next we will consider each of the three different problems, one at a time.

5-3.2 THE FISHER EXACT TEST

As we have outlined for the Fisher "exact" test of "independence", the table total N and the row and column totals may be regarded as being "fixed", and our test of significance should be aimed at judging whether the cell frequencies in the body of the table are reasonable with respect to the row totals, column totals, and table total N, i.e., the inferences therefrom.

It is well-known from combinational theory and elementary probability considerations that the chance of the result depicted in Table 5-2 is

$$\frac{m!n!r!s!}{N!a!b!c!d!} \tag{5-5}$$

which (Ref. 5) may be seen by considering that the contents of the r receptacles marked "defective" form a sample of r from an urn containing m balls marked "Process A" and n balls marked "Process B", the sampling being done without replacement. Hence the probability from Eq. 5-5 added to those of all results less likely than that obtained will form the basis of the Fisher exact test. It can be easily seen, however, that many computations as represented by Eq. 5-5 can result in much drudgery; thus it becomes most desirable to use an approximation for calculating chances.

*Barnard uses CSM in referring to a rather rigorous mathematical ordering of the sample space; the letters mean "convexity, symmetry, and the maximum condition".

**We will refer to this case as the "Double Dichotomy".

Pearson (Ref. 6) points out that the mean value of a in random drawings will be

$$\text{Mean } a = \bar{a} = rm/N \tag{5-6}$$

and that the variance of a is

$$\text{Var}(a) = \sigma_a^2 = mnrs/[N^2(N-1)]. \tag{5-7}$$

One then uses the normal probability tables and Yates' correction for continuity to determine whether the observed result is significant statistically.* This means that the normal probability tables would be entered with

$$z = (a - 0.5 - \bar{a})/\sigma_a = (a - 0.5 - \bar{a})/\ [mnrs/N^2\ (N-1)]^{1/2} \text{ if } a > \bar{a} \tag{5-8}$$

to find the tail area above this observed z, or for the lower tail area use

$$z = (a + 0.5 - \bar{a})/\sigma_a = (a + 0.5 - \bar{a})/\ [mnrs/N^2(N-1)]^{1/2} \text{ if } a < \bar{a} \tag{5-9}$$

as the standardized deviate of entry.

Since contingency table forms, such as Eq. 5-6 or even Eq. 5-5, may be linearized by taking logarithms, much attention has been paid in recent years to "loglinear" models of analysis. See, for example, Section 5-5 on information theory and Section 5-9 of Ref. 7. Also, see Ref. 8. However, for this particular chapter of the handbook, we thought it desirable and had some preference for adhering to analyses of the ***original observations*** for the training of Army statisticians since one would not always use loglinear analyses to the exclusion of the other methods of analysis. In fact, it is very often true that the different methods of analysis give strikingly similar results. Of course, it is also true that with the advent of modern computers and scientific pocket calculators, the matter of transformations to almost any scale of analysis presents no special difficulties.

As a final comment concerning the loglinear models, we quote from the "Introduction" of Fienberg's book (Ref. 7): "The models used throughout this book rely upon a particular approach to the definition of interaction between or among variables in multidimensional contingency tables, based on cross-product ratios of expected cell values. As a result, the models are linear in the logarithms of the expected value scale; hence the label loglinear models.". In connection with the loglinear models and the analysis of cross-classified data falling within the framework of multivariate analyses, it also helps to make a distinction between "response variables", or variables that are free to vary in response to controlled conditions, and "explanatory variables", or variables that are regarded as fixed, either because of the experimentation or because the context of the data suggests they play a determining or causal role.

We continue with the normal approximation, its accuracy, and an example.

Pearson (Ref. 6) discusses the accuracy of the normal approximation for several possible sample sizes of practical interest. Also in some particular cases, one may desire to calculate the Fisher exact probabilities—especially perhaps for low cell numbers—rather than resort to the normal approximation. Indeed, over the years many authorities on statistical methods have advocated that the cell frequencies should be perhaps at least four or five for suitably accurate results from the normal approximation, and sample sizes m and n should be nearly equal.

Instead of the normal approximation there is also the equivalent chi-square approximation. If one squares Eq. 5-8, it can be shown that the result with continuity correction is

$$\chi^2(1) = z^2 = \frac{(|ad - bc| - N/2)^2 N}{mnrs} \quad ** \tag{5-10}$$

*W. G. Cochran has suggested that instead of Yates' correction to "Calculate χ^2 by the usual equation. Find the next lowest possible value of χ^2 to the one to be tested and use the tabular probability for a value of χ^2 midway between the two.".

**The normal method of calculating χ^2 is to take each of the four cell numbers—a, b, c, d—subtract its expected value, square each such difference, then divide by the expected value, and sum the resulting numbers.

and if Yates' continuity correction is not used, chi-square is given as

$$\chi^2 = (ad - bc)^2 N/(mnrs). * \tag{5-11}$$

It should be noted that there is only a single degree of freedom (df) for chi-square.

Perhaps an example of the Fisher exact test would amplify the situation.

Example 5-1:

A class conducted in tank gunnery at Fort Knox had 40 students. The purpose of the class was to teach the students to become expert crew members of the new main battle tank. The instructor also was given the task of selecting the best, or more proficient, gunners for future assignment. The overall program of instruction and training involved not only class study but also actual firing experience in prototype tanks. The instructor noted from student records that 20 of the soldiers had engineering degrees and the others had nonengineering experience. In view of this, it seemed without doubt that the engineers would make the best gunners. Hence the instructor considered that the present class would provide a good "experiment" to test such a hypothesis, and he proceeded to do so. After the complete program of instruction and tank training in the field, nine of the engineers qualified as tank gunners but only six of the nonengineers qualified. Is there sufficient evidence from such a test to conclude that only the engineers should be tank gunners?

We have that $a = 9$, $b = 6$, $c = 11$, $d = 14$, $m = 20$, $n = 20$, $r = 15$, $s = 25$, and $N = 40$. From Eq. 5-6 we find that $\bar{a} = 7.5$, and from Eq. 5-7 we calculate that $\hat{\sigma}_a = 1.55$. Then, by using Eq. 5-8 with the Yates continuity correction to include the observed $a = 9$, we find that the normal deviate $z = 0.65$, which corresponds with an upper tail probability of 0.258** for the normal approximation. Thus assuming that we were conducting the significance test at the 5% level, we would have to conclude that the evidence is not sufficient to say that only engineers make good tank gunners. (We note also in this connection that only 9 of 20 engineers could make the grade.)

Finally, for the Fisher exact test we mention that controversy over the Yates continuity correction continues. Pearson (Ref. 6), on the basis of his many calculations, appears to take the position that the continuity correction is worthwhile for the Fisher exact 2×2 contingency table case although it is very doubtful for comparing binomial populations—the next case discussed. Over the years, many other investigators have tackled the problem of the continuity correction and also have concluded that the Yates correction may well be needed for small frequencies for the Fisher model. Current evidence, therefore, seems to support Yates' continuity correction. The real concern, however, has to do with just how accurately the tail area probabilities have to be determined, including some tie-in with practice, since it may not be too important to distinguish between a probability of 0.05 and 0.07, for example.

5-3.3 THE COMPARISON OF TWO BINOMIAL POPULATIONS

As mentioned earlier, the problem of judging whether two binomial populations have the same parameter p based on small samples selected at random from them probably is one of the more important and frequent activities with which the Army analyst will be concerned. Moreover, there is no really justifiable reason for embedding or hiding this problem in a contingency table; it is important in its own right! Here one selects a sample of m at random from one binomial-type population and also a random sample of size n from a second binomial population and then conducts a significance test to judge whether $p_1 = p_2 = p$, say, where the p's are the true proportions of failures, successes, etc. Many times some product is in service for which the proportion for such a "lot" or population is already fairly well-known; this is treated as a "standard", "control", or best available product. As an example, the best available lot of delay-type antitank fuzes may contain 5% duds. Sometimes this type of lot may be called the "control" lot—a term used in many fields of application—and the true p is designated as p_c. Similarly, once the product is improved (or thought to be improved), the new product to be tested (perhaps for replacing the

*The normal method of calculating χ^2 is to take each of the four cell numbers—a, b, c, d—subtract its expected value, square each such difference, then divide by the expected value, and sum the resulting numbers.

** Pearson (Ref. 6) gives the exact probability for these numbers as 0.2572; therefore, the normal approximation is excellent indeed here. Such cannot be expected for smaller sample sizes or for a very small or very large number of occurrences, however.

"standard" lot), is often referred to as the "test" or "treatment" lot with its proportion of defects, successes, etc., designated as p_t. Thus it is more descriptive to use p_c instead of p_1 and p_t instead of p_2 in practical applications.

For the 2×2 contingency table of Table 5-2, we note that for the comparative binomial case m, n, and N are fixed as before but that the r and s are now random variables, as compared to the Fisher exact test of par. 5-3.2. Thus rather than analyzing the data using all the letters of Table 5-2, we will focus our attention on comparing the sample proportions a/m and b/n, which estimate the true p's. In applying equations, however, it sometimes will be convenient to use the marginals of the table also, especially for comparative purposes with other equations as in par. 5-3.2, for example.

Since we are now dealing with binomial populations, the reader may see that the likelihood of occurrence of the two observed sample results is given by

$$\left[\left(\frac{m!}{a!c!}\right)p_1^a\,(1-p_1)^c\right]\left[\left(\frac{n!}{b!d!}\right)p_2^b\,(1-p_2)^d\right] \tag{5-12}$$

and under the null hypothesis that the two proportions are equal, i.e., $p_1 = p_2 = p$, say, this probability becomes

$$\left(\frac{m!n!}{a!b!c!d!}\right)p^r(1-p)^s. \tag{5-13}$$

One may note that Eq. 5-13 differs from the corresponding likelihood for the Fisher exact test by the factor

$$\left(\frac{N!}{r!s!}\right)p^r(1-p)^s. \tag{5-14}$$

The so-called classical method of testing the null hypothesis that the binomial p's are equal involves taking the estimate $\hat{p}$ of p to be

$$\hat{p} = (a+b)/(m+n) = r/N \tag{5-15}$$

and then using the standard deviation s_d of the difference of the two sample proportions, a/m and b/n, given by

$$s_d = \sqrt{(r/N)\,(1-r/N)N^{-1}\,(1/m+1/n)} \tag{5-16}$$

so that the actual significance test used is the ratio

$$\begin{aligned} difference/s_d &= (a/m - b/n)/\sqrt{(r/N)\,(1-r/N)N^{-1}\,(1/m+1/n)} = z_2, \text{ say } * \\ &= (a - rm/N)/\sqrt{mnrs/N^3}. \end{aligned} \tag{5-17}$$

We note (leaving out the Yates continuity correction), but with some surprise, that Eq. 5-17 is the same as Eqs. 5-8 or 5-9 except that the N^3 of Eq. 5-17 is replaced by the nearly equal factor $N^2\,(N-1)$, which is little different except for small sample sizes! Hence, for even moderate sample size, there is really no difference in the two tests that hypothesize equal p's! However, it is interesting to liken our procedure to the use of Student's t test for comparing two population means in par. 4-7, but especially to the Behrens-Fisher problem of par. 4-7.3.2 for unequal sigmas for the case of continuous variates. Instead of pooling (adding) the numbers of failures or successes from both samples, for example, we might keep them separate and estimate the variances of the population proportions separately. Thus we have

$$\hat{p}_1 = a/m \tag{5-18}$$

* Due to references to the literature, we define Eq. 5-17 as the quantity z_2, and for comparative purposes z_1 is defined next.

and

$$\hat{\sigma}_{\hat{p}_1} = \sqrt{(a/m)\,(1-a/m)/m} \tag{5-19}$$

and an obviously similar quantity for the mean and standard deviation of the estimate b/n of the parameter p_2. Furthermore, we could, as in the Behrens-Fisher problem, and especially granting leeway for the possibility that the standard errors of the proportions are unequal, formulate our significance test as the normal approximation

$$z_1 = (a/m - b/n)/[(a/m)\,(1 - a/m)/m + (b/n)\,(1 - b/n)/n]^{1/2} \tag{5-20}$$

as compared to that of Eq. 5-17, which we have already referred to as z_2. Thus we have competition concerning the better choice for the case of unequal binomial population sigmas between z_1 or Eq. 5-20 as compared to z_2 or Eq. 5-17. This, in fact, is a problem that has recently broken into the statistical literature. Robbins (Ref. 9) points out that when $m = n$, then for equal sample sizes

$$|z_1| \geq |z_2| * \tag{5-21}$$

and asks the important question concerning just which of the two procedures has the better power against possible alternatives to the null hypothesis of equal p's. (It is understood in this connection that the normal approximation calls for about equal p's and that the sample sizes m and n be sufficiently large. Of course, the p's are "about equal" otherwise a statistical test would not be needed, and the sample sizes should be ample for the approximations to hold.) It might be said that there is some advantage in equal sample sizes for then z_1 tends to be larger than z_2 and hence has greater power in the critical region, i.e., a value of z that goes beyond the value 1.96, or the upper 5% point of the standardized normal distribution.

About the time of the Robbins letter to the editor, Eberhardt and Fligner (Ref. 10) had also studied the same question raised by Robbins (Ref. 9) and had arrived at some definite conclusions about the problem. They pointed out that z_1 is in fact more powerful when equal sample sizes are used but that either procedure can be more powerful when sample sizes are unequal. Eberhardt and Fligner (Ref. 10) note also that the test using z_2 is practically equivalent to the chi-square test of Eq. 5-10 or Eq. 5-11 and that Goodman (Ref. 11) had recommended z_1^2 as a competitor to the chi-square test. In addition, if the quantity $(p_1 - p_2)$ were subtracted from the numerator of Eq. 5-20 and the denominator were unchanged, then this sample statistic could be used to advantage to test the hypothesis that $(p_1 - p_2)$ equals some value other than zero. Finally, Lehman (Ref. 12) has shown that, for small samples, the appropriate solution for testing equality of the two binomial p's with a known conditional significance level is in fact the Fisher exact test. Eberhardt and Fligner (Ref. 10) summarize their findings by saying that the "large-sample" comparison favors the test based on z_1 not on z_2 although for small samples there are some contingencies, and we quote: "It was found that for smaller samples the exact size of the test based on z_1 can be much larger than the nominal level, although this is in part compensated for by a corresponding increase in power. For example, when the nominal level is 0.05, the exact size was found to be 0.08075 when $m = n = 20$ and 0.08479 when $m = 40$ and $n = 20$. Thus, the use of z_1 may not be advisable for these smaller sample sizes. However, for the larger sample sizes considered, the exact probabilities tabulated lend some credence to the large-sample comparison.".

Conover (Ref. 13) reiterates that when one is interested in a confidence interval on the true unknown $(p_1 - p_2)$, there is no justification for a pooled variance; therefore, z_1 is preferred because it has more power when $m = n$, but this is the only case for such a result. For all cases of unequal sample sizes, there will be some values of a and b for which the absolute values of the z's will cross each other in size. Conover (Ref. 13) gives an algebraic description of the affected regions. He concludes, "Thus, the choice between z_1 and z_2 for hypothesis testing is inconclusive. Since z_1 is the obvious choice when forming confidence intervals or when testing $p_1 = p_2 + h$ for some specified $h \neq 0$, perhaps z_1 should be selected for the

* Equality occurs only when $a = b$.

case $h = 0$ also.". Presently available research, therefore, still leaves the problem somewhat open, and it seems clear that the practicing analyst might lean toward the use of z_1, which treats the "Behrens-Fisher" type of occurrence. At least, it seems to be more general and "robust".

There are some functions of the unknown parameters p_1 and p_2 about which the practicing statistician may have some special interest in establishing confidence limits. These include the difference Δ, the ratio R, and the odds ratio ψ of p_1 and p_2, which are

$$\Delta = p_1 - p_2 = p_t - p_c \tag{5-22}$$

$$R = p_1/p_2 = p_t/p_c \tag{5-23}$$

and

$$\begin{aligned}\psi &= p_1(1-p_2)/[p_2(1-p_1)] \\ &= p_t(1-p_c)/[p_c(1-p_t)]\end{aligned} \tag{5-24}$$

where the last quantity is well-known as Fisher's odds ratio. The choice of Δ, R, or ψ often is somewhat a matter of personal taste although in many applications the proper choice of one over the other might be clear.

Thomas and Gart (Ref. 14) have published a table of exact confidence limits for the difference, the ratio, and the odds ratio of the unknown p_1 and p_2. The Thomas and Gart (Ref. 14) tables are for the 95% confidence limits and are based on the conditional distribution since Fisher has argued that "the marginal frequencies by themselves supply no information on the point at issue" but that the information they do supply is "wholly ancillary". The relevant conditional distribution used by Thomas and Gart is given as their Eq. 2.1 in Ref. 14. Their 95% confidence limits table is reproduced here as Table 5-3.

Table 5-3 is for equal sample sizes, $m = n$, only, and Lehman (Ref. 12) has pointed out that for this case the best power is obtained for testing the hypothesis that the odds ratio is unity.

To use the table, one calculates a/m and b/n and then labels the smaller of the two ratios as p_c and the larger one as p_t ; these ratios are used to enter Table 5-3. The sample sizes are for only $m = n = 20\ (20)\ 100$, and the P values listed are the one-tail probabilities for the Fisher exact test. Then the lower and upper confidence limits Δ_L and Δ_U, respectively, for the difference of the two p's; the lower and upper 95% confidence limit R_L and R_U, respectively, for the ratio; and then finally lower and upper confidence limits of the odds ratio ψ_L and ψ_U , respectively, are given.

Some questions have been raised concerning whether Table 5-3 gives exact confidence bounds, especially for the difference and the ratio of the two population p's. This point is explained in the corrigenda to the paper (Ref. 14), and we quote Thomas and Gart.

"The limits for the difference, Δ, and the ratio, R, are not exact in the sense that for some values of p_c and p_t the intervals may cover the true values of Δ and R with probabilities, in the conditional sample space, less than the nominal confidence coefficient, $1 - \alpha$. Apparently, this follows from the fact that the marginal total, $a + b = r$, is not an appropriate ancillary statistic to condition on when making inferences on Δ and R. However, for values of p_c and p_t for which $n(p_c + p_t) \approx r$, additional calculations show that the coverage probabilities for Δ and R are similar to those for ψ. The limits for ψ have the coverage probabilities $\geq (1 - \alpha)$ for all p_c and p_t in this conditional sample space. Similarly, all three pairs of limits include the null values ($\psi = R = 1$, $\Delta = 0$, for all p_c and p_t) whenever the exact $p \geq \alpha/2$ and exclude them whenever the exact $P < \alpha/2$."

In summary, since some further investigation may be called for and exact confidence limits for all cases have not appeared in print, the Thomas and Gart table of Ref. 14 will serve as a valuable aid until replaced with a more exact one. Recently, the paper of Santner and Snell (Ref. 15) appeared that proposed three methods of constructing exact confidence bounds. One of the methods should be selected and tables computed to compare with and perhaps replace the Thomas and Gart table of Ref. 14.

With this account of confidence intervals for the various functions of parameters of interest in applications, we find it desirable to make a few remarks about binomial data tests, especially for small sample sizes and about some available tables the Army analyst or statistician might use. Table A-26 of Ref. 2 gives sample sizes required for comparing a proportion with a standard or control proportion when the sign of the difference between the two is important.

TABLE 5-3

EXACT *P* VALUES AND EXACT 95% CONFIDENCE LIMITS FOR DIFFERENCES IN PROPORTIONS IN PERCENT (100Δ), RATIOS OF PROPORTIONS *R*, AND ODDS RATIOS ψ (Ref. 14)

Smaller $100\hat{p}_c$	Larger $100\hat{p}_t$	$m = n$	P value	($100\Delta_L$, $100\Delta_U$)	(R_L, R_U)	(ψ_L, ψ_U)
1*	1*	20	—	(— , —)	(— , —)	(— , —)
1*	1*	40	—	(— , —)	(— , —)	(— , —)
1*	1*	60	0.75	(−3.25, 3.25)	(0.01, 77.22)	(0.01, 79.81)
1*	1*	80	0.75	(−2.44, 2.44)	(0.01, 77.54)	(0.01, 79.47)
1	1*	100	0.75	(−1.95, 1.95)	(0.01, 77.74)	(0.01, 79.28)
1*	2*	20	—	(— , —)	(— , —)	(— , —)
1*	2*	40	0.50	(−2.37, 100.00)	(0.03, ∞)	(0.03, ∞)
1*	2*	60	0.75	(−3.25, 3.25)	(0.01, 77.22)	(0.01, 79.81)
1*	2*	80	0.50	(−3.03, 3.69)	(0.11, 116.52)	(0.10, 120.98)
1	2	100	0.50	(−2.43, 2.95)	(0.11, 116.82)	(0.10, 120.37)
1*	5	20	0.50	(−4.74, 100.00)	(0.03, ∞)	(0.03, ∞)
1*		40	0.25	(−3.37, 100.00)	(0.20, ∞)	(0.19, ∞)
1*		60	0.31	(−4.01, 6.58)	(0.25, 154.82)	(0.24, 165.73)
1*		80	0.18	(−2.63, 6.19)	(0.41, 194.44)	(0.40, 207.27)
1		100	0.11	(−1.63, 5.95)	(0.57, 234.05)	(0.56, 248.86)
1*	10	20	0.24	(−6.62, 100.00)	(0.20, ∞)	(0.19, ∞)
1*		40	0.06	(−1.80, 100.00)	(0.69, ∞)	(0.68, ∞)
1*		60	0.06	(−1.58, 11.58)	(0.76, 271.04)	(0.75, 306.56)
1*		80	0.02	(0.59, 11.19)	(1.11, 350.16)	(1.12, 394.28)
1		100	<.005	(2.08, 10.95)	(1.47, 429.26)	(1.50, 482.05)
1*	25	20	0.02	(0.26, 100.00)	(1.02, ∞)	(1.02, ∞)
1*		40	<.005	(10.23, 100.00)	(2.38, ∞)	(2.68, ∞)
1*		60	<.005	(11.15, 26.58)	(2.44, 617.89)	(2.77, 841.72)
1*		80	<.005	(14.15, 26.19)	(3.34, 815.39)	(3.93, 1104.8)
1		100	<.005	(16.07, 25.95)	(4.24, 1013.1)	(5.10, 1368.2)
5	5	20	0.76	(−9.74, 9.74)	(0.01, 74.53)	(0.01, 82.58)
		40	0.69	(−8.59, 8.59)	(0.08, 13.21)	(0.07, 14.46)
		60	0.66	(−7.56, 7.56)	(0.14, 7.19)	(0.13, 7.79)
		80	0.64	(−6.77, 6.77)	(0.19, 5.19)	(0.18, 5.58)
		100	0.63	(−6.17, 6.17)	(0.24, 4.22)	(0.22, 4.50)
5	10	20	0.50	(−11.94, 14.73)	(0.11, 111.66)	(0.10, 130.99)
		40	0.34	(−7.98, 13.64)	(0.31, 21.10)	(0.28, 24.46)
		60	0.25	(−5.69, 12.67)	(0.45, 11.87)	(0.42, 13.61)
		80	0.18	(−4.23, 11.92)	(0.56, 8.75)	(0.54, 9.96)
		100	0.14	(−3.21, 11.35)	(0.65, 7.21)	(0.63, 8.16)
5	25	20	0.09	(−6.67, 29.73)	(0.64, 221.18)	(0.59, 314.93)
		40	0.01+	(2.22, 28.68)	(1.16, 44.40)	(1.19, 62.43)
		60	<.005	(6.12, 27.75)	(1.51, 25.68)	(1.63, 35.70)
		80	<.005	(8.36, 27.04)	(1.77, 19.27)	(1.96, 26.56)
		100	<.005	(9.84, 26.49)	(1.98, 16.08)	(2.22, 22.02)
5	30	20	0.05−	(−3.17, 34.73)	(0.83, 256.87)	(0.80, 393.83)
		40	<.005	(6.55, 33.68)	(1.46, 52.04)	(1.58, 78.74)
		60	<.005	(10.69, 32.76)	(1.88, 30.22)	(2.14, 45.20)
		80	<.005	(13.04, 32.05)	(2.19, 22.74)	(2.56, 33.70)
		100	<.005	(14.59, 31.50)	(2.43, 19.00)	(2.90, 27.97)
5	40	20	0.01−	(4.89, 44.72)	(1.24, 326.22)	(1.32, 590.83)
		40	<.005	(15.71, 43.68)	(2.07, 67.04)	(2.54, 119.64)
		60	<.005	(20.16, 42.76)	(2.62, 39.17)	(3.41, 69.02)
		80	<.005	(22.65, 42.06)	(3.03, 29.58)	(4.06, 51.61)
		100	<.005	(24.28, 41.51)	(3.34, 24.79)	(4.59, 42.92)
10	10	20	0.70	(−17.05, 17.05)	(0.08, 12.57)	(0.07, 15.21)
		40	0.64	(−13.34, 13.34)	(0.20, 5.01)	(0.17, 5.81)
		60	0.62	(−11.17, 11.17)	(0.28, 3.53)	(0.25, 4.00)
		80	0.60	(−9.76, 9.76)	(0.34, 2.91)	(0.31, 3.24)
		100	0.59	(−8.76, 8.76)	(0.39, 2.56)	(0.35, 2.82)
10	25	20	0.20	(−12.56, 32.14)	(0.47, 23.49)	(0.40, 34.86)
		40	0.07	(−4.01, 28.66)	(0.79, 10.05)	(0.76, 14.27)
		60	0.03	(−0.16, 26.60)	(0.99, 7.34)	(0.99, 10.16)
		80	0.01+	(2.10, 25.24)	(1.13, 6.18)	(1.16, 8.41)
		100	<.005	(3.62, 24.27)	(1.23, 5.52)	(1.29, 7.43)
10	30	20	0.12	(−9.27, 37.14)	(0.62, 27.00)	(0.56, 43.33)
		40	0.02	(0.14, 33.69)	(1.01, 11.68)	(1.01, 17.91)
		60	0.01−	(4.26, 31.65)	(1.24, 8.58)	(1.31, 12.81)
		80	<.005	(6.65, 30.30)	(1.40, 7.24)	(1.52, 10.63)
		100	<.005	(8.25, 29.32)	(1.52, 6.49)	(1.68, 9.41)
10	40	20	0.03	(−1.52, 47.12)	(0.94, 33.73)	(0.92, 64.62)
		40	<.005	(9.06, 43.70)	(1.44, 14.87)	(1.63, 27.09)
		60	<.005	(13.54, 41.67)	(1.74, 11.01)	(2.09, 19.48)
		80	<.005	(16.10, 40.33)	(1.95, 9.35)	(2.42, 16.22)
		100	<.005	(17.79, 39.37)	(2.10, 8.41)	(2.67, 14.39)
10	50	20	0.01−	(7.19, 57.07)	(1.27, 39.90)	(1.41, 94.80)
		40	<.005	(18.49, 53.65)	(1.89, 17.89)	(2.47, 40.13)
		60	<.005	(23.17, 51.64)	(2.26, 13.35)	(3.15, 28.96)
		80	<.005	(25.83, 50.31)	(2.51, 11.38)	(3.65, 24.15)
		100	<.005	(27.58, 49.35)	(2.70, 10.27)	(4.03, 21.45)
20	20	20	0.65	(−25.77, 25.77)	(0.22, 4.62)	(0.16, 6.40)
		40	0.61	(−18.68, 18.68)	(0.36, 2.75)	(0.29, 3.48)
		60	0.59	(−15.26, 15.26)	(0.45, 2.23)	(0.37, 2.71)
		80	0.58	(−13.19, 13.19)	(0.50, 1.98)	(0.43, 2.34)
		100	0.57	(−11.77, 11.77)	(0.55, 1.83)	(0.47, 2.12)
20	25	20	0.50	(−23.35, 30.87)	(0.32, 5.37)	(0.23, 8.04)
		40	0.39	(−15.10, 23.87)	(0.50, 3.26)	(0.41, 4.45)
		60	0.33	(−11.30, 20.48)	(0.60, 2.67)	(0.52, 3.48)
		80	0.29	(−9.02, 18.41)	(0.67, 2.38)	(0.59, 3.03)
		100	0.25	(−7.47, 16.99)	(0.72, 2.21)	(0.65, 2.76)
20	30	20	0.36	(−20.27, 35.91)	(0.42, 6.10)	(0.32, 9.94)
		40	0.22	(−11.14, 28.98)	(0.64, 3.76)	(0.55, 5.56)
		60	0.15	(−7.04, 25.61)	(0.75, 3.10)	(0.69, 4.38)
		80	0.10	(−4.61, 23.55)	(0.83, 2.78)	(0.78, 3.82)
		100	0.07	(−2.97, 22.13)	(0.89, 2.59)	(0.85, 3.48)
20	40	20	0.15	(−12.83, 45.85)	(0.65, 7.48)	(0.54, 14.76)
		40	0.04	(−2.49, 39.03)	(0.92, 4.72)	(0.89, 8.37)
		60	0.01+	(2.01, 35.71)	(1.07, 3.94)	(1.10, 6.64)
		80	<.005	(4.64, 33.66)	(1.17, 3.56)	(1.25, 5.81)
		100	<.005	(6.40, 32.25)	(1.24, 3.32)	(1.36, 5.31)
20	50	20	0.05−	(−4.32, 55.59)	(0.88, 8.71)	(0.83, 21.73)
		40	<.005	(6.79, 48.89)	(1.21, 5.63)	(1.35, 12.42)
		60	<.005	(11.53, 45.62)	(1.39, 4.74)	(1.67, 9.87)
		80	<.005	(14.27, 43.61)	(1.51, 4.30)	(1.89, 8.65)
		100	<.005	(16.10, 42.21)	(1.60, 4.04)	(2.05, 7.92)
25	25	20	0.64	(−28.51, 28.51)	(0.27, 3.65)	(0.19, 5.37)
		40	0.60	(−20.36, 20.36)	(0.42, 2.37)	(0.32, 3.12)
		60	0.58	(−16.57, 16.57)	(0.50, 1.99)	(0.40, 2.49)
		80	0.57	(−14.29, 14.29)	(0.56, 1.80)	(0.46, 2.18)
		100	0.56	(−12.74, 12.74)	(0.59, 1.68)	(0.50, 2.00)
25	30	20	0.50	(−25.47, 33.57)	(0.37, 4.13)	(0.26, 6.62)
		40	0.40	(−16.45, 25.51)	(0.54, 2.73)	(0.43, 3.90)
		60	0.34	(−12.36, 21.74)	(0.63, 2.31)	(0.53, 3.12)
		80	0.30	(−9.93, 19.47)	(0.69, 2.10)	(0.61, 2.75)
		100	0.26	(−8.28, 17.91)	(0.74, 1.97)	(0.66, 2.52)
25	40	20	0.25	(−18.11, 43.48)	(0.56, 5.04)	(0.43, 9.83)
		40	0.12	(−7.87, 35.58)	(0.78, 3.42)	(0.70, 5.87)
		60	0.06	(−3.37, 31.86)	(0.90, 2.92)	(0.86, 4.73)
		80	0.03	(−0.73, 29.61)	(0.98, 2.67)	(0.97, 4.17)
		100	0.02	(1.05, 28.06)	(1.03, 2.52)	(1.05, 3.84)
25	50	20	0.10	(−9.65, 53.11)	(0.77, 5.85)	(0.66, 14.50)
		40	0.02	(1.37, 45.39)	(1.04, 4.07)	(1.06, 8.70)
		60	<.005	(6.11, 41.75)	(1.18, 3.51)	(1.30, 7.03)
		80	<.005	(8.87, 39.54)	(1.27, 3.23)	(1.46, 6.22)
		100	<.005	(10.72, 38.01)	(1.33, 3.06)	(1.58, 5.73)
25	60	20	0.03	(−0.45, 62.37)	(0.99, 6.51)	(0.98, 21.95)
		40	<.005	(11.06, 54.93)	(1.30, 4.65)	(1.58, 13.16)
		60	<.005	(15.95, 51.40)	(1.46, 4.06)	(1.93, 10.62)
		80	<.005	(18.77, 49.25)	(1.57, 3.76)	(2.18, 9.38)
		100	<.005	(20.66, 47.77)	(1.64, 3.57)	(2.36, 8.63)
30	30	20	0.63	(−30.56, 30.56)	(0.33, 3.08)	(0.21, 4.79)
		40	0.60	(−21.63, 21.63)	(0.47, 2.13)	(0.34, 2.90)
		60	0.58	(−17.56, 17.56)	(0.55, 1.83)	(0.43, 2.35)
		80	0.57	(−15.13, 15.13)	(0.60, 1.67)	(0.48, 2.08)
		100	0.56	(−13.47, 13.47)	(0.63, 1.58)	(0.52, 1.92)

* Values of $\hat{p}$ (0.01 and 0.02) where $N\hat{p}$ is not an integer are rounded to the closest integer (0, 1, or 2).

(cont'd on next page)

TABLE 5-3 (cont'd)

Smaller $100p_c$	Larger $100p_t$	$m = n$	P value	$(100\Delta_L, 100\Delta_U)$	(R_L, R_U)	(ψ_L, ψ_U)	Smaller $100p_c$	Larger $100p_t$	$m = n$	P value	$(100\Delta_L, 100\Delta_U)$	(R_L, R_U)	(ψ_L, ψ_U)
30	40	20	0.37	(−23.23, 40.45)	(0.50, 3.74)	(0.35, 7.11)	40	70	20	0.06	(−5.60, 59.05)	(0.90, 3.32)	(0.80, 15.98)
		40	0.24	(−13.09, 31.72)	(0.68, 2.66)	(0.56, 4.37)			40	0.01−	(5.80, 50.94)	(1.11, 2.72)	(1.26, 9.83)
		60	0.17	(−8.61, 27.70)	(0.78, 2.31)	(0.68, 3.56)			60	<.005	(10.67, 47.17)	(1.21, 2.50)	(1.54, 8.01)
		80	0.12	(−5.97, 25.30)	(0.84, 2.13)	(0.77, 3.16)			80	≤.005	(13.51, 44.89)	(1.28, 2.38)	(1.73, 7.11)
		100	0.09	(−4.19, 23.65)	(0.89, 2.02)	(0.83, 2.92)			100	<.005	(15.40, 43.32)	(1.33, 2.30)	(1.87, 6.57)
30	50	20	0.17	(−14.79, 49.98)	(0.69, 4.33)	(0.54, 10.51)	40	80	20	0.01+	(4.90, 65.06)	(1.09, 3.37)	(1.23, 32.72)
		40	0.05+	(−3.88, 41.50)	(0.91, 3.16)	(0.85, 6.49)			40	<.005	(16.50, 58.53)	(1.32, 2.90)	(2.00, 18.74)
		60	0.02	(0.84, 37.58)	(1.02, 2.77)	(1.04, 5.30)			60	<.005	(21.37, 55.34)	(1.43, 2.71)	(2.48, 14.88)
		80	0.01−	(3.61, 35.21)	(1.09, 2.57)	(1.16, 4.71)			80	<.005	(24.18, 53.38)	(1.50, 2.60)	(2.81, 13.04)
		100	<.005	(5.47, 33.60)	(1.15, 2.45)	(1.26, 4.35)			100	<.005	(26.04, 52.01)	(1.55, 2.53)	(3.05, 11.93)
30	60	20	0.06	(−5.60, 59.05)	(0.88, 4.82)	(0.80, 15.98)	50	50	20	0.62	(−33.88, 33.88)	(0.49, 2.02)	(0.24, 4.10)
		40	0.01−	(5.80, 50.94)	(1.14, 3.61)	(1.26, 9.83)			40	0.59	(−23.70, 23.70)	(0.62, 1.62)	(0.38, 2.63)
		60	<.005	(10.67, 47.17)	(1.27, 3.20)	(1.54, 8.01)			60	0.57	(−19.18, 19.18)	(0.68, 1.47)	(0.46, 2.17)
		80	<.005	(13.51, 44.89)	(1.35, 2.99)	(1.73, 7.11)			80	0.56	(−16.50, 16.50)	(0.72, 1.40)	(0.51, 1.95)
		100	<.005	(15.40, 43.32)	(1.41, 2.86)	(1.87, 6.57)			100	0.56	(−14.69, 14.69)	(0.74, 1.34)	(0.55, 1.81)
30	70	20	0.01+	(4.21, 67.34)	(1.09, 5.12)	(1.18, 26.26)	50	60	20	0.38	(−24.62, 42.46)	(0.63, 2.26)	(0.36, 6.29)
		40	<.005	(15.93, 59.87)	(1.38, 3.98)	(1.90, 15.87)			40	0.25	(−14.01, 32.93)	(0.77, 1.85)	(0.57, 3.99)
		60	<.005	(20.87, 56.35)	(1.53, 3.58)	(2.33, 12.83)			60	0.18	(−9.35, 28.64)	(0.84, 1.70)	(0.68, 3.29)
		80	<.005	(23.72, 54.21)	(1.62, 3.37)	(2.63, 11.34)			80	0.13	(−6.61, 26.08)	(0.89, 1.62)	(0.77, 2.94)
		100	<.005	(25.62, 52.73)	(1.69, 3.23)	(2.85, 10.44)			100	0.10	(−4.76, 24.33)	(0.92, 1.57)	(0.83, 2.73)
40	40	20	0.63	(−33.09, 33.09)	(0.41, 2.41)	(0.24, 4.25)	50	70	20	0.17	(−14.79, 49.98)	(0.78, 2.43)	(0.54, 10.51)
		40	0.59	(−23.20, 23.20)	(0.55, 1.82)	(0.37, 2.69)			40	0.05+	(−3.88, 41.50)	(0.94, 2.06)	(0.85, 6.49)
		60	0.57	(−18.79, 18.79)	(0.62, 1.61)	(0.45, 2.21)			60	0.02	(0.84, 37.58)	(1.01, 1.91)	(1.04, 5.30)
		80	0.56	(−16.17, 16.17)	(0.66, 1.51)	(0.51, 1.98)			80	0.01−	(3.61, 35.21)	(1.06, 1.83)	(1.16, 4.71)
		100	0.56	(−14.40, 14.40)	(0.70, 1.44)	(0.55, 1.83)			100	<.005	(5.47, 33.60)	(1.10, 1.78)	(1.26, 4.35)
40	50	20	0.38	(−24.62, 42.46)	(0.57, 2.79)	(0.36, 6.29)	50	80	20	0.05−	(−4.32, 55.59)	(0.94, 2.49)	(0.83, 21.73)
		40	0.25	(−14.01, 32.93)	(0.73, 2.15)	(0.57, 3.99)			40	<.005	(6.79, 48.89)	(1.11, 2.21)	(1.35, 12.42)
		60	0.18	(−9.35, 28.64)	(0.81, 1.93)	(0.68, 3.29)			60	<.005	(11.53, 45.62)	(1.19, 2.08)	(1.67, 9.87)
		80	0.13	(−6.61, 26.08)	(0.86, 1.82)	(0.77, 2.94)			80	<.005	(14.27, 43.61)	(1.25, 2.01)	(1.89, 8.65)
		100	0.10	(−4.76, 24.33)	(0.90, 1.74)	(0.83, 2.73)			100	<.005	(16.10, 42.21)	(1.28, 1.96)	(2.05, 7.92)
40	60	20	0.17	(−15.42, 51.23)	(0.73, 3.10)	(0.54, 9.61)	50	90	20	0.01−	(7.19, 57.07)	(1.11, 2.38)	(1.41, 94.80)
		40	0.06	(−4.33, 42.24)	(0.92, 2.46)	(0.84, 6.06)			40	<.005	(18.49, 53.65)	(1.30, 2.24)	(2.47, 40.13)
		60	0.02	(0.48, 38.14)	(1.01, 2.23)	(1.02, 4.99)			60	<.005	(23.17, 51.64)	(1.40, 2.17)	(3.15, 28.96)
		80	0.01−	(3.29, 35.68)	(1.07, 2.11)	(1.14, 4.45)			80	<.005	(25.83, 50.31)	(1.45, 2.12)	(3.65, 24.15)
		100	<.005	(5.17, 34.00)	(1.11, 2.03)	(1.23, 4.12)			100	<.005	(27.58, 49.35)	(1.49, 2.09)	(4.03, 21.45)

Reprinted with permission. Copyright © by the American Statistical Association.

Table A-28 of Ref. 2 gives what is called "minimum contrasts" for $m = n = 1(1)20(10)100(50)200(100)500$ corresponding to signficance levels of 5% and 1% for a two-sided test on proportions or for a one-sided test with half of these percentage points, i.e., 2.5% and 0.5%. By "minimum contrast" is meant the "least different" pair of observed failures, successes, etc., which is significant at the chosen significance level. A "more different" pair is significant also. For example, if Table A-28 indicates that the pair (1,7) is statistically significant, then so is the pair (1,8), etc.—hence the use of the term "minimum contrast".

Often the Army analyst will have occasion to make a significance test for binomial proportions or for a 2×2 table for small but unequal sample sizes. Here Table A-29 of Ref. 2 would be very valuable because it covers unequal m and n up to 20 and the 0.05, 0.025, 0.01, and 0.001 levels of probability or significance. Also the exact probabilities under the null hypothesis tested are listed in Table A-29 so that such information may be used as an aid to judgment.

We will give two examples, one for small sample sizes and the other for moderate to "large" sample sizes—at least in practice.

Example 5-2:

In a test of 12 "standard" antitank rounds fired at 2000 m against an old tank, all 12 rounds hit the target. Nine experimental antitank rounds were also fired at the same target, but only seven hit the old tank. Can it not be said that the experimental round is much inferior to the standard antiarmor round insofar as hit probability is concerned?

Here we assume that both samples were drawn randomly from the two lots, and we identify that:

$m = 12$, $a = 12$ (hits), $n = 9$ and $b = 7$ (hits).

Note that we are dealing with rather small sample sizes in this problem so we should calculate exact chances or use a table. Because the use of Table A-29 of Ref. 2 is suggested here, we enter that table with (in their notation) $n_1 = 12$ (our m), $n_2 = 9$ (our n), and $a_1 = 12$ (our a). Then for the one-sided 0.05 significance level we find that a statistically significant result for our b (their a_2) would not occur until this number of hits was five or less. Therefore, from this limited test we cannot conclude that the experimental round is inferior in hit probability.

As a matter of calculative interest, we might use Eq. 5-17 and find that from a normal probability table the observed chance or upper tail area is 0.17 with the continuity correction, indicating no significant difference at the 5% level either. We note that Table A-29 of Ref. 2 shows an exact chance of 0.021 in the upper tail area for a_2 = our $b = 5$, the largest value at which significance would occur, but we observed $b = 7$. Had we actually used $b = 5$ instead of 7 and the continuity correction, the normal approximation would have resulted in an upper tail area of 0.022, which shows very excellent agreement with the value of 0.021 from Table A-29 of Ref. 2 and a closeness not expected!

Example 5-3:

Combat simulations, or computerized war games, are often played to represent a given "real" battle time of interest, and the losses on each side are counted to give an indication of the effectiveness of Blue versus Red. One of the major problems concerning future wars centers around always having the best available weapons, and for the infantry this turns out to be rather difficult indeed because major breakthroughs for hand-held weapons are few. Nevertheless, Blue had developed a new rifle sight, a new type of cartridge, and best of all a light machine gun that would not "jump all over the place". To test the effectiveness of his new weapons, the Blue Commander decided to conduct a computerized combat simulation of one of his companies with his new, improved weapon mix against the usual Red company—organized and equipped for such a battle. For the combat situation, the Blue Commander had some special interest in probable results from about 60 of his infantrymen with the newly developed weapons against 60 Red infantrymen in a close skirmish. For the close combat situation played in this connection, there were 18 Red infantrymen lost versus only 6 Blue infantrymen. Since, in the past, Blue and Red infantrymen in such a struggle seemed equally matched, can it be said now that Blue's new weapon mix would show clear superiority? Assume representativeness with future companies.

For this problem we have

$m = n = 60$, $a = 18$, $b = 6$, $c = 42$, $d = 54$, $r = 24$, $s = 96$, and $N = 120$. Also $p_1 = 0.30$ and $p_2 = 0.10$ so that we use $p_c = 0.10$, along with $\hat{p}_t = 0.30$ to enter the Thomas-Gart Table 5-3.

We note from Table 5-3 that the P value is only 0.01 and that the 95% confidence limits on the true differences in p's, the ratio of p's, and the odds ratio are, respectively,

$$(\Delta_L, \Delta_U) = (0.0426, 0.3165)$$
$$(R_L, R_U) = (1.24, 8.58)$$
$$(\psi_L, \psi_U) = (1.31, 12.81).$$

Moreover, it is noted that the 95% confidence limits do not include any of the null values of the parameters, i.e., zero for the difference in the two population p's, or for the ordinary ratio, or for the odds ratio equal to unity. Hence we should very definitely conclude that Blue's new weapon mix is superior to Red's and that it would be expected to inflict 30% Red casualties as compared to only 10% for Blue. (The reader might use the normal approximation of Eq. 5-17 but with the continuity correction to check that the P value so obtained is about 0.006, which agrees with the tabled value of 0.01^-. Alternatively, by noting there seems to be an improvement in Blue's weapons, i.e., 0.30 versus 0.10, and hence there is likelihood of different variances for the contrasts, it seems clear that Eq. 5-20 should be used.)

Thomas and Gart (Ref. 14) also point out that their tables should be used for the planning of experiments. Thus already in possession of good evidence concerning the control proportion, some evidence about the improved process or treatment, and the chosen significance level of 95%, the tables may be

scanned to determine the various P values arising from the use of different sample sizes. For instance, Example 5-3 contains evidence of a Red ability to kill only about 10% of Blue's engaged infantry while Blue, with his new weapon list, is able to kill probably 30% of the Reds. From Table 5-3, therefore, one sees a one-tailed P value of 0.12 for a sample of size 20; then a probability of only 0.02 for $m = n = 40$; a $P = 0.01^-$ for samples of 60 (as we just observed); and when the sample size exceeds 80, the one-tailed probability is less than 0.005. Hence a sample size corresponding to 35 for the chosen significance level of 5% in this case might well be able to detect the indicated difference. Gail and Gart (Ref. 16) discuss the traditional method of selecting the sample size for comparative binomial trials in their 1973 paper. Interested readers may make a comparison of the two methods.

5-3.4 RECENT WORK ON COMPARING TWO BINOMIAL PERCENTAGES

Procedures for comparing two binomial populations, i.e., the comparison of two proportions in 2×2 contingency tables, are fraught with some basic difficulties. In fact, there are continuing arguments on which method of computation should be selected. The Fisher "exact" probabilities, which often have been used for the problem, have been criticized as a "randomization or permutation" test only; the Type I errors or level of significance chosen cannot be guaranteed due to the discrete number of occurrences of "failures" or "successes"; and even though significance test calculations are often nearly the same, there is much difficulty in providing exact confidence bounds on the parameters, or functions of them. Moreover, there is the problem of the continuity correction for the normal and chi-square approximations. To improve the accuracy and practical worth of statistical analyses, Garside (Ref. 17) has published some new continuity correction factors for the chi-square test. By this, we mean that the observed a, b, c, and d are replaced by $a - c_g$, $b + c_g$, $c + c_g$, and $d - c_g$, respectively. For Yates' correction $c_g = 0.5$, but for Garside's correction c_g is a tabulated adjustment depending on m, n, and the significance level α. Boschloo (Ref. 18), in connection with an alternative approach for the smaller sample sizes, has proposed tables of "raised conditional levels of significance" that, if used in place of Fisher's exact test, are still conservative in making judgments but not as much as Fisher's. The Type I error in Fisher's exact test is always less than α, as originally calculated by Fisher; however, many statisticians argue—and perhaps rightly so—that the Type I error for any test should be as close as possible to α, the significance level chosen, and yet not exceed α. Thus although his complete tables have not yet been published, Boschloo's aim is to try to bring the Type I error rates of Fisher's test closer to the nominal level α.

In 1976 Garside and Mack (Ref. 19) carried out some rather extensive calculations to determine error rates for the uncorrected chi-square approximation, Yates' corrected chi-square test, Garside's continuity corrections, Fisher's exact test, and the Boschloo modification of Fisher's test. These computations show that Boschloo's and Garside's error probabilities are very similar, and both are closer to α than either Fisher's or Yates' error rates. Also, as expected, the computations show that the uncorrected chi-square gives probabilities often exceeding the significance level α and that the excess may be appreciable for unequal m and n. Moreover, if α is very small, say 0.001, the error rates may be as much as six times the nominal or expected value α for the uncorrected chi-square test and hence very undesirable.

Concerning randomization tests, Tocher (Ref. 20) proposed a randomization test—which is a modification of Fisher's test—that gives actual Type I error rates exactly equal to α whether none, one, some, or all of the marginal totals are fixed. However, too few users and statisticians really care to make a decision that may depend on the drawing of a random number over and above his observed data! Therefore, such an approach is unlikely to gain confidence for the 2×2 comparative binomial trials contingency tables.

Our discussion so far seems to lead one to conclude that some problems remain concerning the small sample sizes for contingency tables and binomial comparisons. In fact, it is apparently this background of the problem that has led McDonald, Davis, and Milliken (Ref. 21) to propose a nonrandomized, unconditional test for comparing two binomial populations. Indeed, McDonald,

Davis, and Milliken (Ref. 21) have recommended *against* the practice of using the following three tests in comparing binomial p's in the case of small samples: (1) the uniformly most powerful unbiased (UMPU) test of Lehman (Ref. 12) because it depends on randomization, (2) the usual, nonrandomized analogue of the UMPU test, and (3) Fisher's test not only because of its conservativeness but also because there is disagreement with Fisher's philosophy in this case. Thus McDonald, Davis, and Milliken (Ref. 21) find their position more in agreement with that of Barnard (Ref. 5) and Pearson (Ref. 6), and they propose a nonrandomized, unconditional test of the hypothesis $H_0: p_1 = p_2$, which is in the "spirit" of the Barnard-Pearson approach. The McDonald, Davis, and Milliken proposed test is primarily for small sample sizes, $m = n \leq 15$, and they give useful tables of their significance test procedure. It should be pointed out that there are some very desirable features of the McDonald, Davis, and Milliken tables; these are that the exact Type I errors are given and also that the boundaries for the one-sided 5% and 1% (and two-sided) aimed-at or nominal significance levels are included, which may be very helpful.

In the course of their study, McDonald, Davis, and Milliken found that the usually nonrandomized, conditional tests for comparing binomial p's for independent samples are very conservative in the sense that the actual significance level attributable to an outcome is often one-fourth to one-half of the anticipated value. As is well-known, the actual size of the critical region depends on the unknown p for the null hypothesis, or in other words, $\alpha = f(p)$, so that by numerical methods a least upper bound α^* for α can be found, and the actual size of the test must be less than or equal to this least upper bound although the target level may be higher. In their construction of critical regions, McDonald, Davis, and Milliken select a target size, call it α^+, and then for the sum of the observed numbers a and b their critical region consists of those points inside the critical region of the UMPU test. Once the points of such a critical region have been determined, McDonald, Davis, and Milliken assume the independence of a and b to resolve a function $f_1(p_1, p_2)$ for the calculation of the size of the region. The size of the critical region under the null hypothesis then reduces to a function of p, or $f(p)$, which may be studied to find its maximum. If the value of p is some value other than the one that causes $f(p)$ to reach its maximum, the true value of $\alpha = f(p)$ is less than max $f(p)$, and the test is conservative. Therefore, a computer routine is used to evaluate numerically the least upper bound $\alpha^* = \max f(p)$, and finally, a "driver" program iterates on values of α^+ to obtain critical regions with α^* less than or equal to the nominal levels, 5% and 1%, desired. As it turns out, the least upper bound on the size of a two-sided critical region is not necessarily twice the least upper bound on the size of the corresponding one-sided regions. Nevertheless, the sizes of all critical regions are recorded for the sake of judgment.

The McDonald, Davis, and Milliken tables, along with the boundaries of their critical regions and Type I error sizes of Ref. 21, are given here as Table 5-4. To use Table 5-4, the sample size m is taken as less than or equal to the sample size n; then values of the observed number of occurrences a in m observations are listed to the left of the aimed-at one-sided nominal levels of 5% and 1%. The body of each table lists the values of b corresponding to a that will give the boundaries of the critical regions.* The left column within each vertical strip of Table 5-4 gives the upper left critical region boundary values or points for a two-dimensional graph or chart on which a is the abscissa and b is the ordinate. The values listed in the right-hand column of each strip, one for the 5% level and the other for the 1% level, give the lower right-hand boundary points of b for that corner critical region. Note that within each of the two strips of Table 5-4 the sizes of the critical regions are listed; the size for the one-sided test is in the left column, and the size for the two-sided test is in the right column. For the smaller sample sizes, the two-sided size of the critical region is also equal to the one-sided value, and the Type I errors do not approach the desired sizes except for the larger values of m and n. It is somewhat striking that the desired sizes of the critical regions are hardly ever those precisely targeted. Nevertheless, McDonald, Davis, and Milliken's Table 5-4 may prove to be of considerable value in many practical analyses, and we will now illustrate its use.

*There is an upper left critical region and a lower right critical region. Each critical region consists of the boundary and all points (a,b) more distant than expectations under the null hypothesis.

TABLE 5-4

2×2 CONTINGENCY TABLES: TEST FOR COMPARING TWO PROPORTIONS (Ref. 21)
CRITICAL REGIONS FOR THE NONRANDOMIZED UNCONDITIONAL TEST OF H_0: $p_1 = p_2$

			Nominal Level (one-sided)			
			0.05		0.01	
m	n	a				
2	3		*0.035*	*0.063*		
		0	$b = 3$			
		1				
		2		$b = 0$		
2	4		*0.022*	*0.031*		
		0	$b = 4$			
		1				
		2		0		
2	5		*0.015*	*0.016*		
		0	$b = 5$			
		1				
		2		0		
2	6		*0.039*	*0.055*		
		1	$b = 5$			
		2		1		
2	7		*0.030*	*0.032*	*0.009*	*0.009*
		0	$b = 6$		7	
		1				
		2		1		0
2	8		*0.023*	*0.024*	*0.007*	*0.007*
		0	$b = 7$		8	
		1				
		2		1		0
2	9		*0.042*	*0.046*	*0.005*	*0.005*
		0	$b = 7$		9	
		1				
		2		2		0
2	10		*0.034*	*0.035*	*0.004*	*0.004*
		0	$b = 8$		10	
		1				
		2		2		0
2	11		*0.029*	*0.029*	*0.004*	*0.004*
		0	$b = 9$		11	
		1				
		2		2		0
2	12		*0.044*	*0.046*	*0.003*	*0.003*
		0	$b = 9$		12	
		1				
		2		3		0
2	13		*0.038*	*0.038*	*0.009*	*0.009*
		0	$b = 10$		12	
		1				
		2		3		1
2	14		*0.033*	*0.033*	*0.008*	*0.008*
		0	$b = 11$		13	
		1				
		2		3		1
2	15		*0.049*	*0.049*	*0.007*	*0.007*
		0	$b = 11$		14	
		1	15	0		
		2		4		1
3	3		*0.016*	*0.031*		
		0	$b = 3$			
		1				
		2				
		3		0		
3	4		*0.040*	*0.078*	*0.008*	*0.016*
		0	$b = 3$		4	
		1				
		2				
		3		$b = 1$		$b = 0$
3	5		*0.046*	*0.070*	*0.005*	*0.008*
		0	$b = 4$		5	
		1	5			
		2		0		
		3		1		0
3	6		*0.050*	*0.098*	*0.003*	*0.004*
		0	$b = 4$		6	
		1	6			
		2		0		
		3		2		0
3	7		*0.035*	*0.063*	*0.010*	*0.016*
		0	$b = 5$		6	
		1	7			
		2		0		
		3		2		1
3	8		*0.047*	*0.094*	*0.007*	*0.009*
		0	$b = 5$		7	
		1	8			
		2		0		
		3		3		1
3	9		*0.035*	*0.065*	*0.005*	*0.006*
		0	$b = 6$		8	
		1	9			
		2		0		
		3		3		1

(cont'd on next page)

Note: $m \leq n$

TABLE 5-4 (cont'd)

			Nominal Level (one-sided)			
			0.05		0.01	
m	*n*	*a*				
			0.047	*0.051*	*0.004*	*0.004*
3	10	0	*b* = 7		9	
		1	9			
		2		1		
		3		3		1
			0.043	*0.073*	*0.009*	*0.009*
3	11	0	*b* = 7		9	
		1	10			
		2		1		
		3		4		2
			0.036	*0.051*	*0.007*	*0.007*
3	12	0	*b* = 8		10	
		1	11			
		2		1		
		3		4		2
			0.041	*0.074*	*0.009*	*0.009*
3	13	0	*b* = 8		11	
		1	12		13	
		2		1		0
		3		5		2
			0.033	*0.054*	*0.005*	*0.008*
3	14	0	*b* = 9		12	
		1	13		14	
		2		1		0
		3		5		2
			0.048	*0.079*	*0.009*	*0.009*
3	15	0	*b* = 9		12	
		1	13		15	
		2		2		0
		3		6		3
			0.035	*0.070*	*0.004*	*0.008*
4	4	0	*b* = 3		4	
		1	4			
		2				
		3		*b* = 0		
		4		1		*b* = 0
			0.045	*0.078*	*0.002*	*0.004*
4	5	0	*b* = 3		5	
		1	5			
		2				
		3		0		
		4		2		0

			Nominal Level (one-sided)			
			0.05		0.01	
m	*n*	*a*				
			0.026	*0.051*	*0.007*	*0.014*
4	6	0	*b* = 4		5	
		1	6			
		2				
		3		0		
		4		2		1
			0.047	*0.094*	*0.009*	*0.012*
4	7	0	*b* = 4		6	
		1	6		7	
		2				
		3		1		0
		4		3		1
			0.042	*0.066*	*0.006*	*0.007*
4	8	0	*b* = 5		7	
		1	7		8	
		2	8	0		
		3		1		0
		4		3		1
			0.038	*0.074*	*0.008*	*0.012*
4	9	0	*b* = 5		7	
		1	8		9	
		2	9	0		
		3		1		0
		4		4		2
			0.045	*0.075*	*0.005*	*0.007*
4	10	0	*b* = 6		8	
		1	8		10	
		2	10	0		
		3		2		0
		4		4		2
			0.040	*0.079*	*0.008*	*0.014*
4	11	0	*b* = 6		8	
		1	9		11	
		2	11	0		
		3		2		0
		4		5		3
			0.046	*0.086*	*0.010*	*0.011*
4	12	0	*b* = 6		9	
		1	10		11	
		2	12	0		
		3		2		1
		4		6		3
			0.046	*0.086*	*0.008*	*0.008*
4	13	0	*b* = 7		10	
		1	10		12	
		2	13	0		
		3		3		1
		4		6		3

Note: $m \leq n$

(cont'd on next page)

TABLE 5-4 (cont'd)

			Nominal Level (one-sided)			
			0.05		0.01	
m	*n*	*a*				
			0.050	*0.091*	*0.009*	*0.012*
4	14	0	$b = 7$		10	
		1	11		13	
		2	13	1		
		3		3		1
		4		7		4
			0.044	*0.072*	*0.007*	*0.008*
4	15	0	$b = 8$		11	
		1	12		14	
		2	14	1		
		3		3		1
		4		7		4
			0.031	*0.062*	*0.001*	*0.002*
5	5	0	$b = 3$		5	
		1	5			
		2	5			
		3		0		
		4		0		
		5		2		0
			0.049	*0.085*	*0.006*	*0.012*
5	6	0	$b = 3$		5	
		1	5		6	
		2	6			
		3		0		
		4		1		0
		5		3		1
			0.028	*0.056*	*0.008*	*0.017*
5	7	0	$b = 4$		5	
		1	6		7	
		2	7			
		3		0		
		4		1		0
		5		3		2
			0.045	*0.087*	*0.005*	*0.010*
5	8	0	$b = 4$		6	
		1	6		8	
		2	8			
		3		0		
		4		2		0
		5		4		2
			0.043	*0.072*	*0.008*	*0.012*
5	9	0	$b = 5$		7	
		1	7		8	
		2	8			
		3		1		
		4		2		1
		5		4		2
			0.046	*0.068*	*0.010*	*0.015*
5	10	0	$b = 5$		7	
		1	8		9	
		2	9	0	10	
		3	10	1		0
		4		2		1
		5		5		3
			0.044	*0.085*	*0.010*	*0.019*
5	11	0	$b = 5$		7	
		1	8		10	
		2	10	0	11	
		3	11	1		0
		4		3		1
		5		6		4
			0.034	*0.063*	*0.007*	*0.013*
5	12	0	$b = 6$		8	
		1	9		11	
		2	11	0	12	
		3	12	1		0
		4		3		1
		5		6		4
			0.044	*0.087*	*0.009*	*0.012*
5	13	0	$b = 6$		9	
		1	9		11	
		2	12	$b = 0$	13	
		3	13	1		$b = 0$
		4		4		2
		5		7		4
			0.045	*0.086*	*0.008*	*0.015*
5	14	0	$b = 6$		9	
		1	10		12	
		2	12	0	14	
		3	14	2		0
		4		4		2
		5		8		5
			0.047	*0.093*	*0.010*	*0.020*
5	15	0	$b = 7$		9	
		1	10		13	
		2	13	0	15	
		3	15	2		0
		4		5		2
		5		8		6

Note: $m \leq n$

(cont'd on next page)

TABLE 5-4 (cont'd)

			Nominal Level (one-sided)			
			0.05		0.01	
m	*n*	*a*				
			0.034	*0.068*	*0.003*	*0.006*
6	6	0	$b = 3$		5	
		1	5		6	
		2	6			
		3	6	0		
		4		0		
		5		1		0
		6		3		1
			0.048	*0.092*	*0.007*	*0.012*
6	7	0	$b = 4$		5	
		1	5		7	
		2	6		7	
		3	7	0		
		4		1		0
		5		2		0
		6		3		2
			0.034	*0.066*	*0.010*	*0.020*
6	8	0	$b = 4$		5	
		1	6		7	
		2	7		8	
		3	8	0		
		4		1		0
		5		2		1
		6		4		3
			0.044	*0.081*	*0.006*	*0.013*
6	9	0	$b = 4$		6	
		1	6		8	
		2	8		9	
		3	9	0		
		4		1		0
		5		3		1
		6		5		3
			0.045	*0.086*	*0.009*	*0.016*
6	10	0	$b = 4$		7	
		1	7		8	
		2	8		10	
		3	10	0		
		4		2		0
		5		3		2
		6		6		3
			0.049	*0.093*	*0.008*	*0.015*
6	11	0	$b = 5$		7	
		1	7		9	
		2	9		11	
		3	10	1		
		4		2		$b = 0$
		5		4		2
		6		6		4
			0.050	*0.079*	*0.008*	*0.011*
6	12	0	$b = 5$		8	
		1	8		10	
		2	10	0	11	
		3	11	1		
		4	12	2		1
		5		4		2
		6		7		4
			0.043	*0.083*	*0.009*	*0.012*
6	13	0	$b = 6$		8	
		1	8		11	
		2	11	0	12	
		3	12	1	13	0
		4	13	2		1
		5		5		2
		6		7		5
			0.042	*0.078*	*0.009*	*0.018*
6	14	0	$b = 6$		8	
		1	9		11	
		2	11	0	13	
		3	13	1	14	0
		4	14	3		1
		5		5		3
		6		8		6
			0.047	*0.092*	*0.007*	*0.013*
6	15	0	$b = 6$		9	
		1	9		12	
		2	12	0	14	
		3	14	1	15	0
		4	15	3		1
		5		6		3
		6		9		6
			0.038	*0.075*	*0.006*	*0.013*
7	7	0	$b = 3$		5	
		1	5		6	
		2	6		7	
		3	7	0		
		4	7	0		
		5		1		0
		6		2		1
		7		4		2

Note: $m \leq n$

(cont'd on next page)

TABLE 5-4 (cont'd)

			Nominal Level (one-sided)			
			0.05		0.01	
m	*n*	*a*				
			0.037	*0.065*	*0.008*	*0.014*
7	8	0	*b* = 4		5	
		1	5		7	
		2	7		8	
		3	8	0	8	
		4	8	0		0
		5		1		0
		6		3		1
		7		4		3
			0.045	*0.081*	*0.008*	*0.016*
7	9	0	*b* = 4		6	
		1	6		7	
		2	7		9	
		3	8	0	9	
		4	9	1		0
		5		2		0
		6		3		2
		7		5		3
			0.043	*0.079*	*0.009*	*0.016*
7	10	0	*b* = 4		6	
		1	6		8	
		2	8		9	
		3	9	0	10	
		4	10	1		0
		5		2		1
		6		4		2
		7		6		4
			0.042	*0.085*	*0.006*	*0.010*
7	11	0	*b* = 4		7	
		1	7		9	
		2	8		10	
		3	10	0	11	
		4	11	1		0
		5		3		1
		6		4		2
		7		7		4
			0.047	*0.090*	*0.008*	*0.015*
7	12	0	*b* = 5		7	
		1	7		9	
		2	9		11	
		3	10	0	12	
		4	12	2		0
		5		3		1
		6		5		3
		7		7		5

			Nominal Level (one-sided)			
			0.05		0.01	
m	*n*	*a*				
			0.049	*0.081*	*0.009*	*0.017*
7	13	0	*b* = 5		7	
		1	8		10	
		2	10		11	
		3	11	1	13	
		4	12	2		0
		5		3		2
		6		5		3
		7		8		6
			0.046	*0.091*	*0.009*	*0.018*
7	14	0	*b* = 5		8	
		1	8		10	
		2	10		12	
		3	12	1	14	
		4	13	2		0
		5		4		2
		6		6		4
		7		9		6
			0.048	*0.090*	*0.009*	*0.016*
7	15	0	*b* = 6		8	
		1	8		11	
		2	11	0	13	
		3	13	1	14	
		4	14	2		1
		5	15	4		2
		6		7		4
		7		9		7
			0.041	*0.082*	*0.004*	*0.008*
8	8	0	*b* = 3		5	
		1	5		7	
		2	6		8	
		3	7	0	8	
		4	8	0		
		5	8	1		0
		6		2		0
		7		3		1
		8		5		3
			0.040	*0.089*	*0.010*	*0.017*
8	9	0	*b* = 4		5	
		1	5		7	
		2	7		8	
		3	8	0	9	
		4	9	0	9	0
		5	9	1		0
		6		2		1
		7		4		2
		8		5		4

Note: $m \leq n$

(cont'd on next page)

TABLE 5-4 (cont'd)

			Nominal Level (one-sided)			
			0.05		0.01	
m	*n*	*a*				
			0.038	*0.073*	*0.009*	*0.017*
8	10	0	*b* = 4		6	
		1	6		7	
		2	7		9	
		3	9	0	10	
		4	10	0	10	0
		5	10	1		0
		6		3		1
		7		4		3
		8		6		4
			0.042	*0.085*	*0.009*	*0.019*
8	11	0	*b* = 4		6	
		1	6		8	
		2	8		9	
		3	9	0	11	
		4	10	1	11	0
		5	11	2		0
		6		3		2
		7		5		3
		8		7		5
			0.042	*0.084*	*0.009*	*0.017*
8	12	0	*b* = 4		6	
		1	7		9	
		2	8		10	
		3	10	0	11	
		4	11	1	12	0
		5	12	2		1
		6		4		2
		7		5		3
		8		8		6
			0.046	*0.088*	*0.008*	*0.016*
8	13	0	*b* = 5		7	
		1	7		9	
		2	9		11	
		3	10	0	12	
		4	12	1	13	0
		5	13	3		1
		6		4		2
		7		6		4
		8		8		6
			0.046	*0.091*	*0.009*	*0.017*
8	14	0	*b* = 5		7	
		1	8		10	
		2	9		11	
		3	11	0	13	
		4	13	1	14	0
		5	14	3		1
		6		5		3
		7		6		4
		8		9		7
			0.041	*0.081*	*0.009*	*0.017*
8	15	0	*b* = 5		8	
		1	8		10	
		2	10		12	
		3	12	0	14	
		4	13	2	15	0
		5	15	3		1
		6		5		3
		7		7		5
		8		10		7
			0.049	*0.098*	*0.010*	*0.021*
9	9	0	*b* = 3		5	
		1	5		6	
		2	6		8	
		3	7	0	8	
		4	8	0	9	
		5	9	1		0
		6	9	2		1
		7		3		1
		8		4		3
		9		6		4
			0.042	*0.076*	*0.010*	*0.020*
9	10	0	*b* = 4		5	
		1	5		7	
		2	7		8	
		3	8	0	9	
		4	9	0	10	
		5	10	1		0
		6	10	2		1
		7		3		2
		8		5		3
		9		6		5
			0.050	*0.100*	*0.010*	*0.018*
9	11	0	*b* = 4		6	
		1	6		7	
		2	7		9	
		2	8	0	10	
		4	10	0	11	
		5	11	1		0
		6	11	3		1
		7		4		2
		8		5		4
		9		7		5

(cont'd on next page)

Note: $m \leq n$

TABLE 5-4 (cont'd)

			Nominal Level (one-sided)			
			0.05		0.01	
m	*n*	*a*				
			0.039	*0.078*	*0.010*	*0.021*
9	12	0	*b* = 5		6	
		1	6		8	
		2	8		9	
		3	9	0	11	
		4	11	0	12	0
		5	12	1	12	0
		6	12	3		1
		7		4		3
		8		6		4
		9		7		6
			0.042	*0.084*	*0.010*	*0.017*
9	13	0	*b* = 4		7	
		1	7		9	
		2	8		10	
		3	10	0	11	
		4	11	1	13	0
		5	12	2	13	0
		6	13	3		2
		7		5		3
		8		6		4
		9		9		6
			0.044	*0.088*	*0.010*	*0.017*
9	14	0	*b* = 5		7	
		1	7		9	
		2	9		11	
		3	10	0	12	
		4	12	1	13	0
		5	13	2	14	1
		6	14	4		2
		7		5		3
		8		7		5
		9		9		7
			0.050	*0.099*	*0.009*	*0.018*
9	15	0	*b* = 5		7	
		1	7		10	
		2	9		11	
		3	11	0	13	
		4	12	1	14	0
		5	14	3	15	1
		6	15	4		2
		7		6		4
		8		8		5
		9		10		8
			0.045	*0.089*	*0.006*	*0.013*
10	10	0	*b* = 4		5	
		1	5		7	
		2	6		8	
		3	8		9	
		4	8	0	10	
		5	9	1	10	0
		6	10	2		0
		7		2		1
		8		4		2
		9		5		3
		10		6		5
			0.044	*0.084*	*0.010*	*0.020*
10	11	0	*b* = 4		6	
		1	5		7	
		2	7		8	
		3	8	0	10	
		4	9	0	10	
		5	10	1	11	0
		6	11	2		1
		7	11	3		1
		8		4		3
		9		6		4
		10		7		5
			0.041	*0.075*	*0.009*	*0.017*
10	12	0	*b* = 4		6	
		1	6		8	
		2	7		9	
		3	9	0	10	
		4	10	0	11	
		5	11	1	12	0
		6	12	2		1
		7	12	3		2
		8		5		3
		9		6		4
		10		8		6
			0.048	*0.093*	*0.008*	*0.015*
10	13	0	*b* = 4		6	
		1	6		8	
		2	8		10	
		3	9	0	11	
		4	10	1	12	0
		5	12	1	13	0
		6	12	3	13	1
		7	13	4		2
		8		5		3
		9		7		5
		10		9		7

Note: $m \leq n$

(cont'd on next page)

TABLE 5-4 (cont'd)

			Nominal Level (one-sided)			
			0.05		0.01	
m	*n*	*a*				
			0.048	*0.085*	*0.008*	*0.016*
10	14	0	$b = 4$		6	
		1	7		9	
		2	8		10	
		3	10	0	12	
		4	11	1	13	0
		5	12	2	14	0
		6	13	3	14	1
		7	14	4		2
		8		6		4
		9		7		5
		10		10		8
			0.043	*0.087*	*0.008*	*0.017*
10	15	0	$b = 5$		7	
		1	7		9	
		2	9		11	
		3	10	0	12	
		4	12	1	14	0
		5	13	2	15	0
		6	14	3	15	1
		7	15	5		3
		8		6		4
		9		8		6
		10		10		8
			0.047	*0.093*	*0.009*	*0.017*
11	11	0	$b = 4$		5	
		1	5		7	
		2	6		8	
		3	8		9	
		4	9	0	10	
		5	9	1	11	0
		6	10	2	11	0
		7	11	2		1
		8		3		2
		9		5		3
		10		6		4
		11		7		6
			0.048	*0.095*	*0.009*	*0.019*
11	12	0	$b = 4$		5	
		1	5		7	
		2	7		9	
		3	8	0	10	
		4	9	0	11	
		5	10	1	11	0
		6	11	2	12	1
		7	12	3		1
		8	12	4		2
		9		5		3
		10		7		5
		11		8		7

			Nominal Level (one-sided)			
			0.05		0.01	
m	*n*	*a*				
			0.043	*0.079*	*0.010*	*0.020*
11	13	0	$b = 4$		6	
		1	6		8	
		2	7		9	
		3	9	0	10	
		4	10	0	12	
		5	11	1	12	0
		6	12	2	13	1
		7	13	3		1
		8	13	4		3
		9		6		4
		10		7		5
		11		9		7
			0.049	*0.097*	*0.010*	*0.018*
11	14	0	$b = 4$		6	
		1	6		8	
		2	8		10	
		3	9	0	11	
		4	10	0	12	0
		5	12	1	13	0
		6	13	2	14	1
		7	13	4	14	2
		8	14	5		3
		9		6		4
		10		8		6
		11		10		8
			0.045	*0.087*	*0.008*	*0.016*
11	15	0	$b = 5$		6	
		1	7		9	
		2	8		10	
		3	10	0	12	
		4	11	1	13	0
		5	12	1	14	0
		6	14	3	15	1
		7	14	4	15	2
		8	15	5		3
		9		7		5
		10		8		6
		11		10		9

Note: $m \leq n$

(cont'd on next page)

TABLE 5-4 (cont'd)

			Nominal Level (one-sided)			
			0.05		0.01	
m	*n*	*a*				
			0.050	*0.099*	*0.009*	*0.017*
12	12	0	$b = 4$		5	
		1	5		7	
		2	6		8	
		3	8		10	
		4	9	0	10	
		5	10	1	11	0
		6	10	2	12	0
		7	11	2	12	1
		8	12	3		2
		9		4		2
		10		6		4
		11		7		5
		12		8		7
			0.048	*0.094*	*0.010*	*0.017*
12	13	0	$b = 4$		5	
		1	6		7	
		2	7		9	
		3	8		10	
		4	9	0	11	
		5	11	1	12	0
		6	11	2	13	0
		7	12	2	13	1
		8	13	4		2
		9		5		3
		10		6		4
		11		7		6
		12		9		8
			0.044	*0.086*	*0.008*	*0.015*
12	14	0	$b = 4$		6	
		1	6		8	
		2	7		9	
		3	9	0	11	
		4	10	0	12	
		5	11	1	13	0
		6	12	2	13	1
		7	13	3	14	1
		8	14	4		2
		9	14	5		3
		10		7		5
		11		8		6
		12		10		8
			0.050	*0.099*	*0.010*	*0.019*
12	15	0	$b = 4$		6	
		1	6		8	
		2	8		10	
		3	9	0	11	
		4	10	0	12	
		5	12	1	14	0
		6	13	2	14	1
		7	14	3	15	1
		8	15	5		3
		9	15	6		4
		10		7		5
		11		9		7
		12		11		9
			0.038	*0.017*	*0.010*	*0.019*
13	13	0	$b = 4$		5	
		1	5		7	
		2	7		8	
		3	8		10	
		4	9	0	11	
		5	10	1	11	0
		6	11	1	12	0
		7	12	2	13	1
		8	12	3	13	2
		9	13	4		2
		10		5		3
		11		6		5
		12		8		6
		13		9		8
			0.050	*0.094*	*0.010*	*0.020*
13	14	0	$b = 4$		6	
		1	6		7	
		2	7		9	
		3	8		10	
		4	9	0	11	
		5	11	1	12	0
		6	12	2	13	0
		7	12	2	14	1
		8	13	3	14	2
		9	14	5		3
		10		6		4
		11		7		5
		12		8		7
		13		10		8

Note: $m \leq n$

(cont'd on next page)

TABLE 5-4 (cont'd)

			Nominal Level (one-sided)			
			0.05		0.01	
m	*n*	*a*				
			0.048	*0.093*	*0.009*	*0.017*
13	15	0	$b = 4$		6	
		1	6		8	
		2	7		9	
		3	9	0	11	
		4	10	0	12	
		5	11	1	13	0
		6	12	2	14	1
		7	13	3	14	1
		8	14	4	15	2
		9	15	5		3
		10	15	6		4
		11		8		6
		12		9		7
		13		11		9
			0.044	*0.088*	*0.008*	*0.017*
14	14	0	$b = 4$		5	
		1	5		7	
		2	7		9	
		3	8		10	
		4	9	0	11	
		5	10	1	12	0
		6	11	1	13	0
		7	12	2	13	1
		8	13	3	14	1
		9	13	4	14	2
		10	14	5		3
		11		6		4
		12		7		5
		13		9		7
		14		10		9
			0.044	*0.085*	*0.010*	*0.019*
14	15	0	$b = 4$		6	
		1	6		8	
		2	7		9	
		3	8		10	
		4	10	0	12	
		5	11	1	12	0
		6	12	2	13	0
		7	13	2	14	1
		8	13	3	15	2
		9	14	4	15	3
		10	15	5		3
		11		7		5
		12		8		6
		13		9		7
		14		11		9
			0.049	*0.099*	*0.009*	*0.018*
15	15	0	$b = 4$		5	
		1	5		7	
		2	7		9	
		3	8		10	
		4	9	0	11	
		5	10	1	12	0
		6	11	1	13	0
		7	12	2	14	1
		8	13	3	14	1
		9	14	4	15	2
		10	14	5	15	3
		11	15	6		4
		12		7		5
		13		8		6
		14		10		8
		15		11		10

Note: $m \leq n$

Reprinted with permission. Copyright © by the American Statistical Association.

Example 5-4:

Since Example 5-2 dealt with sample sizes of only 12 and 9 standard and experimental rounds, respectively, we will use the same data to check that analysis with McDonald, Davis, and Milliken's table.

We have $m = 9$, the smaller sample size, $a = 7$, $n = 12$, and $b = 12$. Then, entering Table 5-4 with $m = 9$, $n = 12$, and $a = 7$, one finds that for the one-sided test at the 0.039 level no b value is given; therefore, again we cannot conclude that the experimental round is really inferior in hit probability to the standard round. We do note, however, that the point $a = 6$ and $b = 12$ is a boundary point on the critical region of this test, but that significance does not result from the test using Table A-22 of Ref. 2 for Example 5-2 unless the value of $a = 5$ were attained. Hence the McDonald, Davis, and Milliken test would reach significance more quickly. Perhaps this shows that a very complete investigation of the critical regions for the binomial and contingency table problems is very worthwhile since it is seen that some rather critical decisions may be necessary.

5-3.5 THE DOUBLE DICHOTOMY

Finally, for the 2×2 contingency table we arrive at what Barnard (Ref. 5) refers to as his third type of abstract experiment and what Pearson (Ref. 6) labels Problem III for the 2×2 contingency table. We will refer to it in this chapter as the "double dichotomy". For the case of the double dichotomy, the sampling is such that a preselected number N of items is drawn from a large population, or universe, and random samples of sizes m and n are obtained, which contain a defective units from the first designated process and b defective units from the second. In this final case for the 2×2 contingency table, we note that only the total sample size N is fixed or preselected, whereas the four row and column totals—m, n, r, and s—are random numbers just as the cell numbers a, b, c, and d are. In this case the multinomial expansion applies, and it seems advantageous and lucid in presentation to use the notation of Pearson (Ref. 6) as a basis for describing the experiment so involved. In fact, Pearson (Ref. 6) describes this case as a test for the independence of two characters, or characteristics, A and B, say. It is supposed that some individuals selected at random will possess character A, while others will not; for this reason we designate them as $\bar{A}$ or "not A". Likewise, some of the individuals in the sample will possess character B and others will not; consequently, we designate them as $\bar{B}$ or "not B". Continuing, let us use the notation $p(A)$ to designate the chance that an individual selected at random will have character A and $p(\bar{A}) = 1 - p(A)$ to designate the probability that such an individual will not possess character A. It is clear then that corresponding probabilities for the character B are $p(B)$ and $p(\bar{B}) = 1 - p(B)$. Finally, we see that four alternative combinations of characteristics will occur: AB, $A\bar{B}$, $\bar{A}B$, and $\bar{A}\bar{B}$. The probabilities for these occurrences are best presented as indicated in Table 5-5.

TABLE 5-5
PROBABILITIES

	A	$\bar{A}$	Total
B	$p(AB)$	$p(\bar{A}B)$	$p(B)$
$\bar{B}$	$p(A\bar{B})$	$p(\bar{A}\bar{B})$	$p(\bar{B})$
Total	$p(A)$	$p(\bar{A})$	1

In terms of the observed sample data, Table 5-6 is shown.

TABLE 5-6
DOUBLE DICHOTOMY TABLE

	A	$\bar{A}$	Total
B	a	c	m
$\bar{B}$	b	d	n
Total	r	s	N

If the null hypothesis specifying the independence of A and B is true, it follows that

$$\begin{aligned} p(AB) &= p(A)p(B), \qquad & p(A\bar{B}) &= p(A)p(\bar{B}) \\ p(\bar{A}B) &= p(\bar{A})p(B), \quad \text{and} \quad & p(\bar{A}\bar{B}) &= p(\bar{A})p(\bar{B}). \end{aligned} \tag{5-25}$$

Hence, we see that given a random sample of size N, the observed data of Table 5-6, which is in the form of a contingency table, may be analyzed to test the hypothesis of independence of the characteristics A and B. As Pearson points out in this case, there is only one application of a random process—i.e., the selection of the total of N individuals, each one of which must fall into one of the four categories of Table 5-6. Furthermore, if another sample of N items were drawn at random, the values of a, b, c, and d would change in a random manner as would the row totals m and n, and the column totals r and s.

The test of independence for the double dichotomy amounts to the test of a composite hypothesis, and the reader easily may see that the probability P of the particular observed result given in Table 5-6 is

$$P = \left(\frac{N!}{a!b!c!d!}\right) p(AB)^a p(A\bar{B})^b p(\bar{A}B)^c p(\bar{A}\bar{B})^d \tag{5-26}$$

which, under the assumption that the null hypothesis of Eq. 5-25 is true, becomes

$$P = \left(\frac{N!}{a!b!c!d!}\right) p(A)^{a+b} p(B)^{a+c} p(\bar{A})^{c+d} p(\bar{B})^{b+d}. \tag{5-27}$$

Furthermore, as Pearson indicates in Ref. 6, Eqs. 5-26 and 5-27 can also be expressed as

$$\begin{aligned} P &= \left(\frac{N!}{m!n!}\right) p(B)^m [1 - p(B)]^n \times \left(\frac{N!}{r!s!}\right) p(A)^r [1 - p(A)]^s \times \frac{m!n!r!s!}{a!b!c!d!N!} \\ &= P_2[m|p(B),N] \times P_2[r|p(A),N] \times P_1[a|N,r,m], \text{ say,} \end{aligned} \tag{5-28}$$

where P_1 and P_2 are specific probabilities.

Refer to Eq. 5-28; Pearson (Ref. 6) points out there are three major factors involved. The first P_2 represents the chance of obtaining in N random observations exactly m items with character B as in Table 5-6, while the second P_2 alike represents the chance of obtaining exactly r items that possess character A of Table 5-6. These two factors, therefore, represent binomial trials as we discussed in par. 5-3.3; there are, however, some differences in notation, of course. Finally, the third, or last, major factor P_1 is precisely the probability of Eq. 5-5 for the Fisher exact test. This third factor specifies that given m items with character B and r items with character A, the chance for the observed partition a, b, c, and d is exactly P_1 or Eq. 5-5. We see, therefore, that the double dichotomy problem involves some of the characteristics of both the Fisher exact test and the comparative binomial trials experiment, especially if we were to test $p(A) = p(B)$.

With regard to a statistical test of significance for the double dichotomy case, one could calculate the probabilities as indicated in Eq. 5-28 for hypothesized values of $p(A)$ and $p(B)$, equal or not, although this would be laborious indeed. However, most often the $p(A)$ and $p(B)$ have to be estimated from the same sample data, and tables to cover many significance tests would be too voluminous. Finally, it seems clear that one must rely on the normal or equivalent chi-square approximation as the obvious choice, i.e., either Eq. 5-8 or Eq. 5-17. In this connection, Pearson (Ref. 6) indicates that the normal approximation along with the continuity correction will be very much on the safe side, i.e., the formal or stated size of the critical region is likely to be much above the actual level attained no matter what the values of $p(A)$ or $p(B)$ are. In fact, the presence of the two binomial terms in Eq. 5-28 will make it likely that overestimation of α will be greater in the double dichotomy problem than in the comparative binomial trials. Thus it would be expected that unless m, n, r, and s are too small, Eq. 5-8 will be a suitable approximation.

In summary, we see that fortunately or unfortunately we are stuck with the normal approximation to a great extent! Nevertheless, there remains much research to be done for the 2×2 contingency table, and

many of the problems encountered also extend to higher order contingency tables, which have even more complications and involvement. With the present state of the art, some readers may not be impressed with the differences we have made concerning essentially three distinct sampling and analytical problems for the 2×2 contingency since the result is just about the same method of analysis except for the smaller sample sizes, in which case we must carry out direct calculations or refer the observed data to an appropriate table. Nevertheless, it certainly seems wise to point out that such distinctions may be important as general guidelines even though it is at the same time fortunate that rather simple normal approximations ordinarily will suffice in many practical applications. Finally, the 2×2 table and higher order tables may be analyzed by using the chi-square principle of summing the squares of deviations from expectations divided by expected values.

Example 5-5 illustrates the principle of the "double dichotomy".

Example 5-5:

In a random sample of 40 recruits at an Army induction and training center, 18 had previous experience with shooting a rifle and 22 did not. Of the 18, 12 of the recruits qualified as "expert"; the other six did not. On the other hand, of the 22 with no former rifle training, 9 trainees qualified as "expert". Can it be said there is conclusive evidence that background experience in shooting rifles is necessary for a trainee to become expert?

This particular example meets the strict requirements for a "double dichotomy" in that both row and column totals, or all marginal totals, can be treated as random variables, and the sample size of 40 is preselected for the experiment to be conducted. Thus by treating the problem this way, we have $N = 40$, $m = 18$, $a = 12$, $n = 22$, $b = 9$, $r = 21$; and $s = 19$. Moreover, the sample size is not small nor are the cell frequencies unusually low. Hence, we may as well use the normal approximation for our analysis. By using Eqs. 5-6 and 5-7 and then by computing z from Eq. 5-8, we have

$$\text{Mean } a = 9.45,\ \sigma_a = 1.591, \text{ and } z = 1.29.$$

This value of z, from a table of the normal integral, corresponds with an upper tail area of about 0.10. Thus it cannot be concluded that background experience in shooting a rifle substantially benefited the recruits because they learned very quickly anyway.

5-3.6 INDEPENDENCE AND INTERACTION IN 2×2 CONTINGENCY TABLES

At this point, it is important to discuss briefly the relation between the concepts of independence and interaction in 2×2 contingency tables. In a two-way classification in the ANOVA for continuous variates, the concept of interaction was perhaps more easily understood, and the reader saw that the interaction term—when there existed only a single observation per cell—was used as the experimental error to judge row and column effects by using an "F" test. On the other hand, for the 2×2 contingency table the concept of interaction is perhaps more difficult to grasp. Independence of association between the cross-classifications in a 2×2 table was defined in terms of the basic probability laws indicating independence in Eq. 5-25. Bartlett (Ref. 22) has defined the meaning of "interaction" as it applies to contingency tables and has stated: "The testing of independence in a 2×2 table may be regarded as testing the significance of the interaction between the two classifications.". Thus insofar as 2×2 contingency tables are concerned, the concepts of independence and interaction are to be taken as being synonymous for all intents and purposes. Therefore, the significance test carried out for a 2×2 contingency table is a test for independence, or a lack of association between the cross-classifications, or a test of the nonexistence of any interaction between the two classifications. This leads us to the use of information theory in the analysis of 2×2 tables as our next pertinent topic.

5-4 SOME DEFINITIONS OF SYMBOLS FOR GENERAL CONTINGENCY TABLES

Before proceeding to the use of Kullback's (Ref. 23) principle of minimum discrimination information estimation analysis of contingency tables or to higher order (multidimensional) contingency tables, it is best to adopt a more general notation than we have used for the 2×2 tables—a notation that was convenient for the purpose of referring to some of the specific and basic papers on the subject. First, suppose we consider only dual classifications again but expand this to the possibility of a two-way table that now has $r \geq 2$ rows and $c \geq 2$ columns. Then we further define:

$x(ij)$ = observed frequency for the cell in the ith row and jth column, for $i = 1, \ldots, r$ and $j = 1, \ldots, c$

$x(i.)$ = sum of the $x(ij)$ across the c columns of the ith row

$x(.j)$ = sum of the $x(ij)$ across the r rows of the jth column

$x(..)$ = N, or sometimes n, = the sum of all observations within the contingency table

$p(ij)$ = true but unknown probability of occurrence, or population proportion, for an individual belonging to the cell in the ith row and jth column of the table

$p(i.)$ = $pr(x = i)$ = marginal probability for ith row

$p(.j)$ = $pr(x = j)$ = marginal probability for jth column.

With these definitions, it is seen, for example, that:

$x(11)$ = observed number of occurrences a for the cross-classification involving A and B in Table 5-6

$x(21)$ = observed number of occurrences given by b for the cross-classification $\bar{B}$ and A as in Table 5-6.

Finally, we will define

$x^*(ij)$ = predicted value for the cell in the ith row and jth column, which is determined in accordance with Kullback's (Ref. 23) minimum discrimination information statistic (MDIS), as discussed in par. 5-5.

Probabilities $p^*(ij)$, $p^*(i.)$ and $p^*(.j)$ may be correspondingly used.

5-5 THE KULLBACK MINIMUM DISCRIMINATION INFORMATION STATISTICS

For a background on the relation of information theory and statistics, the interested readers should study Kullback's *Information Theory and Statistics* (Ref. 24), which covers the basic principles. Perhaps one of the most prominent applications of information theory in statistics has been that concerning the analysis of multidimensional contingency tables by Kullback, and a very useful and readable account of the methodology is that contained in Ref. 23. It is suggested that Army analysts also study Refs. 25, 26, and 27 because they will help to round out the general use of information theory applied to contingency tables.

Kullback's information theory approach to the analysis of contingency tables proceeds basically as follows. First, for any observed contingency table of interest, it seems appropriate to visualize three associated tables:

1. The so-called π table, containing cell elements $\pi(ij)$. The π table may be specified by the null hypothesis, estimated, or given by the observations. For example, the π table may specify the condition or hypothesis of equal probability in all the cells, or it may specify two-way independence, or three-way independence, etc.

2. The second associated table is a p table denoted by the unknown quantities $p(ij)$ defined in par. 5-4. This p table is a contingency table that satisfies certain conditions of interest—for instance, the one-way marginals $p(i.)$, $p(.j)$, etc.

3. The third and final associated table is called the p^* table; the elements of which are denoted by $p^*(ij)$. The p^* table is that member of the class of p tables that most closely resembles the π table in the sense of Kullback's minimum discrimination information, i.e., the π table minimizes the discrimination information given by the equation

$$I(p{:}\pi) = \Sigma p \ln (p/\pi) \tag{5-29}$$

over the class of p tables, where $I(p{:}\pi)$ stands for "information".

Although we can neither go extensively into the details of the Kullback approach, nor is it necessary, we will summarize one or two main results of the information theory approach that are quite germain and very useful insofar as this chapter on contingency tables is concerned. Briefly and for example, we give the following items of some special interest.

If we set

$$\pi\,(ij) = 1/(rc) \tag{5-30}$$

which is the condition for a uniform table with r rows and c columns, the classical hypotheses of independence, homogeneity, conditional independence, no interaction, etc., are represented by p^* tables when certain of the marginals are considered fixed and can be considered as generalized independence hypotheses. The term "generalized independence" means that the cell probability of a multidimensional contingency table may be expressed as the product of factors that are functions of the pertinent marginals.* The more common notions of independence, conditional independence, homogeneity, or conditional homogeneity in contingency tables are all rather special cases of "generalized independence". As Kullback points out in Ref. 24, this is the consequence of the fact that the minimum discrimination information estimates are formulated as members of an exponential family that for the contingency tables application also may be expressed as a multiplicative model or logarithmic linear additive model. Such models are derived on the basis of minimizing the discrimination information. For further appreciation and deeper understanding, interested readers should study Ref. 24 in general and Ref. 26 for the applications to multidimensional contingency tables. The details are, in fact, rather involved. In Ref. 27 Kullback gives a further description of the principles of minimum discrimination information statistics and also presents a 3×2×3×2 example of contingency table analysis, which applies to the firing of guns. Ref. 28 is an earlier paper on the background theory and analysis of contingency tables using the MDIS approach, and Ref. 28 covers the use of loglinear models in the analysis of contingency tables. An excellent and ever-continuing valuable review of contingency tables is available in Ref. 29. We give Kastenbaum's references in our bibliography.

In Ref. 28 Kullback, Kupperman, and Ku summarize some of the more basic principles of the minimum discrimination information statistics and indicate the simplest form of the appropriate estimate of twice the amount of information in terms of observed and expected frequencies and the relation to the well-known chi-square statistic. Quite generally, if we consider, say, r observed frequencies, the ith designated by O_i with $i = 1, \ldots, r$, and E_i defined to be the expected ith frequency (which will be determined with marginal values or totals), the relationship for a one-way contingency table, so to speak, is

$$2\hat{I} = 2\sum_{1}^{r} O_i \ln(O_i/E_i) \approx \Sigma\,(O_i - E_i)^2/E_i = \chi^2\,(r-1) \tag{5-31}$$

that is to say twice the estimate of the amount of information is asymptotically distributed as the chi-square statistic with $(r-1)$ df. Note that twice the estimate of the amount of information is approximately distributed as chi-square, but not exactly. Thus Kullback, Kupperman, and Ku (Ref. 28) show that for contingency tables or "categorical" type data, the minimum discrimination information in its simplest form amounts to summing the observed frequencies multiplied by the natural logarithms of the ratios of the observed to the expected frequencies and to multiplying this result by two; the final expression gives twice the amount of information as is shown in Eq. 5-31. For two-way contingency tables this means that we calculate the double summation given by

$$2\hat{I} = 2\sum_{1}^{r}\sum_{1}^{c} [x\,(ij)] \ln\,[x(ij)/(n\hat{p}_{ij})] = \Sigma\Sigma[x(ij)]\, n \ln[x(ij)]/(x_{i.}x_{.j}) \approx \chi^2\,[(r-1)\,(c-1)] \tag{5-32}$$

in which we have used the appropriate marginals to estimate the unknown p_{ij} . The quantity of Eq. 5-32 for a general number of rows and columns represents the interaction term of the contingency table, which is used to test for independence of row and column effects and is approximately distributed as chi-square with $(r-1)\,(c-1)$ df. Thus for the simple 2×2 table there is only a single df. In Example 5-6 we will apply Eq. 5-32 to the data of Example 5-5.

*As it turns out, most analyses will involve the prediction of cell frequencies from the marginal totals and will not hypothesize a "uniform" table based on Eq. 5-30.

Example 5-6:

Use the data as given in Example 5-5 and apply Kullback's minimum discrimination information theory to determine whether independence exists for the row and column effects, i.e., previous training is not necessary to become an expert. By referring to the symbols of Table 5-6 and the observed quantities of Example 5-5, we have, for example, that the observed number of occurrences $x(12) = c = 6$, whereas the MDIS estimate of this cell value would be $x^*(12)\, x_{1.} x_{.2}/n = (18)\,(19)\,/\,(40) = 8.55$. Proceeding in a like manner, one calculates by Eq. 5-31 for the 2×2 table:

$$2I(x;x^*) = 2[12\ln(12/9.45) + 6\ln(6/8.55) + 9\ln(9/11.55) + 13\ln(13/10.45)] = 2(1.33) = 2.66 \approx \chi^2(1).$$

The approximate upper tail area for an observed chi-square of 2.66 with 1 df is about 0.12; hence we cannot conclude that dependence has been established between rows and columns, i.e., it is necessary to have had extensive training as a rifleman to become an "expert" in the Army training program of rifle shooting.

If interested, one may calculate the ordinary chi-square for the 2×2 table by summing the observed minus the expected values squared divided by the expected values to obtain an observed chi-square of 2.63, which is a little different but nearly the same as that obtained from the information theory approach. This is caused by the use of two different methods, and one therefore should expect small differences in values. These differences will be inconsequential insofar as any judgment is concerned.

Since we have mentioned the matter of a one-way contingency table and to show the generality of Kullback's information theory approach, we give an example from Ref. 28 on tossing coins in Example 5-7.

Example 5-7:

Five coins are thrown in a series of 74 independent tosses, and the number of heads is recorded. We desire to test the hypothesis of a binomial distribution with parameter 1/2, or the chance of a head occurring is 1/2; independence of the trials is assumed. For convenience, the results of the 74 tosses of five coins are brought together in Table 5-7, in which we have calculated and included the expected frequencies. Use the information theory analysis of contingency tables to accept or reject the null hypothesis of a binomial distribution with parameter 1/2 as being the appropriate model to fit the observed data.

TABLE 5-7

SEVENTY-FOUR TOSSES OF FIVE COINS (Ref. 28)

Number of Heads	Theoretical Probability	Observed Frequency	Expected Frequency
0	1/32	2	2.31
1	5/32	5	11.56
2	10/32	22	23.13
3	10/32	29	23.13
4	5/32	14	11.56
5	1/32	2	2.31
		Total = 74	

By using the second expression of Eq. 5-31, we calculate that the observed chi-square is 6.74, and the df are 6 − 1 = 5. Using a table of the percentage points of chi-square, we find that the observed value of 6.74 for 5 df will be exceeded with a probability of about 0.25 and hence is not significant. Thus we accept the null hypothesis of a binomial distribution with parameter of $p = 1/2$ for the chance of tossing a head.

With this rather brief account of the information theory approach to the analysis of contingency tables and other statistical problems, the reader should be impressed with the power and general usefulness of this approach in solving Army problems. Ref. 28 is highly recommended reading and study for Army analysts because it extends Kullback's information theory approach to two-way, three-way, and higher order contingency tables. Moreover, Ref. 28 gives a number of informative examples of applications. We will return to the further use of information theory in connection with the analysis of two-way contingency tables with r rows and c columns in par. 5-8 after some relevant discussion about 2×2 tables.

5-6 SOME RELATED TOPICS AND THE POWER OF 2×2 CONTINGENCY TABLES

In connection with a rather important problem concerning the selection of the "better" of two binomial populations, i.e., for example, the one with the smaller proportion of defectives, Berry and Sobel (Ref. 30) have suggested an "improved" procedure. Their recommendations are based on a "play-the-winner" sampling procedure for determining the better of the two Bernoulli populations under consideration. The procedure these authors have developed is designed to select the better population (call it number 1 with "success" parameter p_1) and with probability P, whenever the difference between the two binomial population parameters ($p_1 - p_2$) is greater than or equal to a specified value Δ, where the quantities P and Δ are preassigned constants. The truncation procedure used by Berry and Sobel is designed to minimize both the expected total number of trials and also the number of trials for one of the populations, i.e., number 2. Moreover, Berry and Sobel's procedure is designed with special reference to the problem of small p's, an important problem in practice. Hence this technique may have application to a number of Army problems.

Darroch (Ref. 31) has discussed the concepts of "multiplicative" and "additive" interaction in contingency tables—the 2×2 table is a special case. The multiplicative and additive definitions of no interaction are compared according to whether they possess or fail to possess the properties of being partitionable, closest to independence, implied by independence, or of placing no constraints on the marginal totals. Further research in this area may be needed to establish the superiority of either the multiplicative or additive interaction concept.

In Ref. 32 Mantel and Hankey update the concept of odds ratios related to 2×2 contingency tables, especially in terms of the three models of interest we have discussed in par. 5-3. Gart (Ref. 33) discusses both point and interval estimates of the odds ratio in the combination of 2×2 tables, and Copas (Ref. 34) gives an account of randomization models for the matched and the unmatched 2×2 tables.

An important topic for practical considerations to which we have barely alluded is that of the power function of 2×2 contingency tables and the related area of determination of proper sample size. With regard to this topic, Casagrande, Pike, and Smith (Ref. 35) give the power function of the exact test for comparing two binomial distributions and include tables of the exact sample sizes ($n = m$) required to test a variety of values for the population proportions p_1 and p_2 ($p_1 > p_2$), at the one-sided significance levels 0.05, 0.025, 0.01, and 0.005, and power requirements of 80%, 90%, and 95%. Their tables also may be used to calculate sample sizes for two-sided tests by entering the tables with half the desired significance level.

5-7 THE GENERAL TWO-WAY CONTINGENCY TABLE (r Rows and c Columns)

5-7.1 INTRODUCTORY FORMULATION

The general two-way contingency table involves a distribution of frequencies in a second order matrix of cells for which there are two or more rows and columns. Our primary desire in this connection is to analyze the observed cross-classification of frequencies to determine whether the row and column effects are independent or unassociated, so to speak. Also for the two-way table we may desire to test for the possible existence of homogeneity and interaction effects as discussed in the sequel. It will be helpful to present the tabular form of frequencies as in Table 5-8.

For the $r \times c$ contingency table one might look at the overall table as a "total" variance based on ($rc - 1$) df from which the row effects based on ($r - 1$) df—or the column effects based on ($c - 1$) df—may be subtracted to give the "conditional" term of fewer total rows (or fewer total columns), and then subtracting the column effects (or row effects) finally gives the interaction or "independence" effect, which is used to judge whether dependence of the cross-classifications does indeed exist. In other words, one might visualize the entire analysis as a two-way ANOVA problem.

In the sequel we will analyze the observed data of Table 5-8 from two points of view. The first approach will be the classical chi-square, where, as previously stated, we sum the squares of the observed minus the expected frequencies divided by, or corrected by, the expected values for each cell. The second will be Kullback's information theory approach. We will also make a comparison of the two by means of an example.

TABLE 5-8

THE GENERAL TWO-WAY CONTINGENCY TABLE

	Column Effects					
	I	II	III	···	c	Totals
Row Effects						
A	$x(11)$	$x(12)$	$x(13)$	···	$x(1c)$	$x(1.)$
B	$x(21)$	$x(22)$	$x(23)$	···	$x(2c)$	$x(2.)$
·						
·						
·						
r	$x(r1)$	$x(r2)$	$x(r3)$	···	$x(rc)$	$x(r.)$
Totals	$x(.1)$	$x(.2)$	$x(.3)$	···	$x(.c)$	$x(..)$ $= N$ or n

5-7.2 THE CLASSICAL CHI-SQUARE ANALYSIS OF TWO-WAY CONTINGENCY TABLES

It is well-known from the statistical literature, or the reader may find it in any standard textbook on statistical methods, that the classical test of independence between the two characteristics specified in a two-way table is based on the statistic

$$\chi^2[(r-1)\,(c-1)] = 2\sum_{1}^{r}\sum_{1}^{c}[x(ij) - x(\mathrm{i}.)x(.j)/N]^2/[x(i.)x(.j)/N] \tag{5-33}$$

where the expected cell frequencies are estimated from the product of appropriate row and column totals divided by the table total. Eq. 5-33 clearly measures the overall amount of the deviations from expectations on a scale consistent with the usual chi-square statistic that is distributed with $(r-1)\,(c-1)$ df, i.e., the same df as for the normal interaction term in the ANOVA. A number of examples of the application of chi-square associated with contingency tables are given in Ref. 1 although we will give a rather special example (Example 5-8), which is quite subjective in nature and, therefore, could raise a number of questions concerning its real validity.

Example 5-8:

In research and development work the Army instituted the practice of a series of in-process reviews (IPR's) for many of its major development programs. The IPR's were considered an effective means of aiding the development process—a good way to assess correctly the status of a project and to bring about effective command coordination. To study the overall effectiveness of IPR's, a series of questionnaires was prepared and distributed to all of the Department of Army (DA) participants for completion. The questionnaires were answered by managers, supervisors, engineers, policymakers, and technicians to obtain a wide spectrum of opinions. The questions were only four in number for a particular phase of the survey—i.e., whether the IPR's were "good in theory and good in practice", "good in theory but poor in practice", "poor in theory but still good in practice nevertheless", or finally "poor in theory and poor in practice both". The whole study with accompanying analyses is covered by Bell, Mioduski, and Belbot in Ref. 36. However, we discuss only a particular contingency table, which is a summary of the results of the IPR Questionnaire. The results we will analyze by using the classical chi-square method for contingency tables are given in our Table 5-9.

An examination of Table 5-9 raises any number of questions about the questionnaire itself! For example, are the categories sufficiently mutually exclusive for a good survey? Are not some or many of the managers really supervisors, and are not many of the policymakers either managers or supervisors, so that the choice of major functions leaves much to be desired? Perhaps the four categories of answers are sufficiently distinct to judge whether the IPR's are really worthwhile although another possible problem is evident, which revolves around whether the answers can be absolutely objective! Indeed, are not the

respondents more or less motivated to answer only two of the rows, namely, that IPR's are good in theory or principle whether good or poor in practice? The last two rows of Table 5-9 are so sparse that the survey appears to encourage only "bureaucratic" answers! Nevertheless, we will look into whether or not independence or association exists between the four forms of answers on one hand and the major function or type of position of the respondents on the other. For this purpose one calculates the chi-square of Eq. 5-33 using the proper row and column totals for expected cell frequencies. In Table 5-9 we have listed the expected cell frequencies in parentheses; an example is the expected value in the second row and third column, i.e., (61)(27)/(137) = 12.0. The calculated value of chi-square is

$$\chi^2(12) = 10.86$$

whereas the upper 5% point of chi-square for 12 df = 21.0. Hence our judgment is that the answers given and the major functions or jobs are not associated, i.e., are independent. This further means that, whether or not the table of data and the conditions under which the data were taken might appear suspicious, we are not able to document conclusively that the lack of objectivity is established. Thus IPR's must be worthwhile.

TABLE 5-9

SUMMARY OF RESULTS OF IPR QUESTIONNAIRE
(By major function)

Answer	Manager	Supervisor	Engineer	Policymaker	Technician	Totals
	(Number responding as indicated by cells)					
Good in Theory and Good in Practice	45 (39.4)*	12 (11.3)	10 (14.6)	1 (3.8)	6 (4.9)	74
Good in Theory and Poor in Practice	27 (32.5)	9 (9.4)	16 (12.0)	6 (3.1)	3 (4.0)	61
Poor in Theory and Good in Practice	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0
Poor in Theory and Poor in Practice	1 (1.1)	0 (0.3)	1 (0.4)	0 (0.1)	0 (0.1)	2
Totals	73	21	27	7	9	137

*The numbers in parentheses are expected values.

As an afterthought, the reader will notice that the last two rows of Table 5-9 appear to be superfluous and really give no information whatever. Moreover, some readers would question our not using the Yates continuity correction. With regard to the latter point, we could refer our calculated value of chi-square to a table of the cumulative probability integral of chi-square for 12 df and use Cochran's recommendation, which places the observed chance at the 53% or 54% level. Therefore, the observed chi-square, with or without the continuity correction, is so far from the 95% point of 21 that adjustment hardly seems worthwhile.

With regard to the lack of positive responses in the last two rows of Table 5-9, we urge the reader, as an exercise, to use only the first two rows of data and to obtain new column totals and a new table total. He may calculate the newly observed chi-square and then draw his own conclusions.

Finally, for the so-called classical method of chi-square analysis, we come to an interesting point. We have used only 12 df; since there were 19 df originally, what about the others? It is easy to see in this connection that the number of df for rows amounts to 3 df, and for columns it is 4 df, which accounts for all 19 df. We could make a further analysis of the row and column variations although it seems to be of little interest, and we will explore this problem further in the sequel by using Kullback's information theory approach to analyze two-way tables.

5-7.3 KULLBACK'S INFORMATION THEORY ANALYSIS OF TWO-WAY TABLES

In contrast to the classical procedure for analyzing the data of a two-way contingency table, we will apply the principles of Kullback's MDIS approach to the problem, i.e., an extension of the analysis in par. 5-7 for 2×2 tables. Refs. 26 and 28 are cited as two of the major pertinent articles concerning the $r \times c$ contingency tables.

In connection with the information theory approach for analyzing $r \times c$ tables, we should keep in mind that originally in par. 5-5 we started with "probability" tables of π, p, and p^* and converted the analysis to deal with the observed numbers, i.e., $x(ij)$ and the marginal predictions $x^*(ij)$. In fact, we used the MDIS

$$2nI(p{:}p^*) = 2I(x{:}x^*) = 2\Sigma\Sigma x(ij)\ln[x(ij)/x^*(ij)] \tag{5-34}$$

which is distributed asymptotically as chi-square under the null hypothesis with an appropriate number of df depending on its composition of terms. Recall also that the minimum discrimination information statistics are used to measure the closeness of resemblance of one table to another since this is their basis for analysis of independence, no interaction, or association, etc. We see, therefore, that for a $r \times c$ table the $x^*(ij)$ are determined from the appropriate marginals just as they were for the classical procedure because this provides the minimum discrimination information, and then Eq. 5-34 may be used to test for independence of row and column effects.

The MDIS of Eq. 5-34 is quite a general one indeed since it extends to a three-way table involving $x(ijk)$, or to a four-way table with frequencies $x(ijkl)$, and to many-way contingency tables as well. Thus Kullback's approach applies to many, many different Army problems involving contingency table analysis.

For the general many-way contingency tables, the MDIS's have a very important property—the Pythagorean property—which is very useful in analysis. The simplest form of the Pythagorean property is expressible in a fundamental theorem by Kullback, which states that

$$2I(x{:}x_1^*) = 2I(x_2^*{:}x_1^*) + 2I(x{:}x_1^*) \tag{5-35}$$

where the subscript 1 refers to a set H_1 of given marginals and the x_2^* corresponds to a set H_2 of given marginals, which is included in the set H_1, which will be illustrated in the discussion that follows. The fundamental theorem states that the MDIS under consideration can be divided into two parts; one is referred to as the measure of effect term represented by the first term on the right-hand side (RHS) of Eq. 5-35, and the second is referred to as a goodness of fit term represented by the second term of the RHS of Eq. 5-35, which results in a test for the existence of any interaction effects. Moreover, this unique Pythagorean property also extends generally to contingency tables of any order (Refs. 26 and 28). This means that a very general analysis is available through the Kullback information theory approach, which proceeds from first testing the significance of one-way marginal effects and a first-order interaction to a test of two-way marginal effects and a second-order interaction, etc., and finally to the highest order interaction term. For two-way contingency tables either the row total effects, using a hypothesized value of $p_{i\cdot}$, or the column total effects, using a hypothesized value of $p_{\cdot j}$, could be tested for significance, and in practice they would most often be significant. Then one would proceed to test the first-order conditional; and finally by subtracting the two-way marginal variations and first-order interaction, one usually would obtain the second-order interaction as the primary test of independence. It is important to know the appropriate number of df of chi-square at each stage or for each test. Unless one is familiar with this approach, he will want to study Refs. 26 and 28. For the two-way table the determination of the number of df is relatively straightforward—there are $(r - 1)$ df for the rows; $(c - 1)$ df for the columns; $r(c - 1)$ df for the first conditional interaction when row variations are first subtracted from the total or $c(r - 1)$ df for the case in which the column variations are first subtracted from the initial total; and finally when row, column, and conditional information is subtracted from the initial total, one reaches the

residual interaction or independence test based on $(r-1)(c-1)$ df. The entire process and a routine for the general analysis are best illustrated by means of an "analysis of information" table, or ANOVA, along with a suitable example (Example 5-9).

Example 5-9:

The Army invited competitive proposals from three of the better machine gun (MG) manufacturers for a new, lighter weight MG to replace the current standard MG. We designate the competing manufacturers by A, B, and C and the standard MG by S. To select the best MG, it was decided with the advice of high Army officials that pop-up targets in a simulated combat environment would be fired upon with each competitive MG, and the number of hits or "kills" recorded. For the experiment each manufacturer would produce 10 prototypes, from which one MG would be randomly selected to compete with a current standard MG also randomly selected from available MG's. All four competitive MG's would be fired randomly by one of the Army's top machine gunners. As a secondary part of the experiment, each MG would be fired until a stoppage of some kind occurred, in which case, and as another issue, the reliability of operation would be evaluated even though the primary focus in the simulated combat environment test is on the analysis of the proportion of hits. The final results of the experiment are brought together in Table 5-10, which is a 2×4 contingency table to be analyzed to test the hypothesis that the proportion of hits is independent of the different MG's.

TABLE 5-10

RESULTS OF MG FIRING EXPERIMENT

	Machine Gun Identification				
	A	*B*	*C*	*S*	Total
Results					
Number of Hits	31	35	24	7	97
Number of Misses	14	41	43	6	104
Total	45	76	67	13	201

Table 5-10 indicates that a stoppage for manufacturer A's machine gun occurred at 45 rounds fired, that his MG obtained 31 hits in the 45 rounds fired, etc., and finally that the current standard MG gave 7 hits in 13 rounds fired before a stoppage occurred. The number of rounds to a stoppage could be analyzed as a reliability evaluation by using the methods of Chapter 21 of Ref. 3, for example. Although for the purposes of a contingency table study or analysis, we will study only the number of hits and misses. Also as a more complex type of problem, one might consider truncating the experiment at points of stoppage in a more refined analysis. In any event, we will view the problem only as a two-way contingency table to illustrate whether the different MG's do in fact show dependence concerning the number, or proportion, of hits and misses.

As a preliminary view of the contingency table experiment, we might expect some differences between the MG's of the different manufacturers—especially since they are possibly competing for a production contract for the best weapon. Moreover, as far as the two rows are concerned, these involve only the number of hits and misses and thereby require no special analysis of such a variation. Thus it seems clear that one would be concerned primarily with the analysis of the interaction term. Nevertheless, all pertinent and ancillary information is brought together in Table 5-11, which includes the general equations for the computation of information for two-way contingency tables. (Chapter equation numbers are at the RHS.)

Note that for Example 5-9 we have calculated the numerical value of the information for only the final interaction term, i.e., the independence test of Eq. 5-40.* As contrasted to the observed chi-square value of 12.34, one may calculate the observed value of the classical chi-square statistic according to, for example, the next to last factor of Eq. 5-31 for the whole table of Table 5-10. If this were done, one would

*Eq. 5-40 is in Table 5-11 on p. 5-42.

TABLE 5-11

TWO-WAY ANALYSIS OF INFORMATION TABLE

Component Due to	Information	df	Calculation of Information $2\hat{I}$	df	
Total	$2\Sigma\Sigma x(ij) \ln\{x(ij)/[np(ij)]\}$	$rc-1$		7	(5-36)
Rows	$2\Sigma\Sigma x(i.) \ln\{x(i.)/[np(i.)]\}$	$r-1$		1	(5-37)
Conditional (Total less rows)	$2\Sigma\Sigma x(ij) \ln\left[\frac{x(ij)\,p(i.)}{x(i.)\,p(ij)}\right]$	$r(c-1)$		6	(5-38)
Columns	$2\Sigma\Sigma x(.j) \ln\{x(.j)/[np(.j)]\}$	$c-1$		3	(5-39)
Independence (Conditional less columns)	$2\Sigma\Sigma x(ij) \ln\left[\frac{nx(ij)}{x(i.)\,x(.j)}\right]$	$(r-1)(c-1)$	12.34	3	(5-40)

find that the observed classical chi-square would be 12.13. By way of comparison, the two calculations of chi-square by the two different approaches to the analysis of contingency tables are about equal, as one might expect, and therefore, serve as a check. Moreover, the observed value of chi-square, i.e., of $2\hat{I}$, with 3 df is highly significant because the corresponding probability is about 0.007. Consequently, we conclude that the results of the test are highly dependent on the MG manufacturers and, in particular, that different manufacturers' weapons will give different true hit probabilities. It appears, therefore, that manufacturer A may have the higher hit probability, i.e., about $31/45 = 0.69$, and that the standard weapon has a hit probability of perhaps as low as 0.54 in addition to fewer rounds to a stoppage. It also could be that manufacturer A may have a problem in reliability as compared to manufacturers B and C since the number of rounds to a stoppage is lower, i.e., 45 versus 76 and 67, respectively. We will not go into an analysis of rounds to stoppage, i.e., the reliability, any further in this example. (See Ref. 3.)

Finally, we return to the numerical calculation of information for the total table based on $(rc - 1)$ df, the row variations, the first conditional term, and the columns. For these terms, or Eqs. 5-36 through 5-39, the true unknown probabilities $p(ij)$, $p(i.)$, and $p(.j)$ are needed. If sound values for the these parameters were available from the physical aspects of the problem, they could be used and all information calculations could be made. However, since this is not the case, we have tested for significance only for the independence, or final interaction, term of Eq. 5-40 since no information beyond that available from the experiment is required. We also remark in this connection that it does not seem at all desirable to hypothesize that the individual cell probabilities $p(ij)$ should be taken to be $1/(nrc)$, or, that is, the condition of a uniform distribution of hits. In fact, there is a much more logical procedure by which to obtain theoretical frequencies for this particular test or experiment, especially if one has appropriate data on the performance of the weapons from past or ancillary tests. By knowing the delivery accuracy of the weapons and the target size and shape, one could calculate the probabilities of hitting and use these as the basis for the $p(ij)$, $p(i.)$, and $p(.j)$. Thus many experiments could exist for which appropriate theoretical cell probabilities may be determined; however, on the other hand, there also will be many, many cases in which only the observed data of the experiment at hand can or should be used. In summary, care should always be exercised in determination of appropriate theoretical frequencies if such information is to be used to draw sound inferences. Finally, one sees the desirability of planning the experiment beforehand to be sure not only that the best or appropriate observations are taken, but also that any possible important physical theories or conditions are tested for significance.

Having covered two-way contingency tables, we now direct our attention to three-way and higher order contingency table analysis procedures.

5-8 COMMENTS ON THE ANALYSIS OF THREE-WAY AND HIGHER ORDER CONTINGENCY TABLES

The principles we have discussed so far in this chapter for the statistical analysis of 2×2 and two-way contingency tables extend, but often with some difficulty, to three-way and higher order tables. In fact, both the classical chi-square approach and Kullback's information theory approach may be used, and even as a check, for the higher order tables. Even though both techniques result in the final use of a chi-square computation, the reader may see that the information theory approach does indeed appear to handle problems of interest in a more elegant manner than the classical chi-square developments. An excellent reference for the analysis of higher order contingency tables using the information theory approach is that of Kullback, Kupperman, and Ku (Ref. 28). This reference gives the theory and several illustrative examples for the two-way and the higher order contingency tables, which the Army analyst may follow and use to advantage in his work. The classical chi-square approach to similar problems is covered and documented in the references and especially in the bibliography. Hence we will conclude this introductory account of modern analyses of contingency tables by citing some typical examples that the reader may find of some value in his applications of higher order tables.

With reference to some of the recent, typical Army applications, it seems that the US Army Operational Test and Evaluation Agency, with Kullback as a consultant, has made an extensive number of applications of the information theory approach to experiments involving operational test type data for Army personnel and equipment. For example, Withers (Ref. 37) cites the use of the Kullback fundamental theorem, or the Pythagorean relation, with applications to several particular operational test and evaluation programs. One covered an operational test of the DRAGON antiarmor weapon and involved the use of a 3×2×2 contingency table. A principal point of inquiry concerning this test was the selection of the best of three different training programs to produce DRAGON gunners capable of engaging both stationary and moving targets. In this operational test 108 missile firings by three groups of 36 gunners were arranged into a 3×2×2 contingency table for analysis. It was learned that although the three different training procedures had significant effects, the target mode, i.e., stationary or moving, had larger effects on hit probability. Moreover, some quantitative information on the relative importance of both the training procedures and the target mode was extracted from the experimental data.

Another operational experiment involved the squad automatic weapon, and the statistical analysis is discussed by Withers (Ref. 37). The purpose of this test was to determine the operational effectiveness of three different types of squad automatic weapons. In a subtest of the overall experiment, 40 silhouette targets depicting enemy fire teams of squad size at four different ranges were randomly presented for engagement. The response variable for this test program and analysis involved the percent of targets hit. The total amount of data represented over 9000 firings, of which 263 targets were hit out of 1804 engagements, and was arranged into a four-way contingency table for analysis. As a result, there were insufficient data to show, even for such a large sample, that the three different types of squad automatic weapons had any effects on target hitting capability. Target range and weapon burst size effects were also analyzed and are reported in Ref. 37.

Another operational experiment for which a contingency table analysis was advantageous involved two candidate target location radars for artillery; these data are also reported in Ref. 37. In this case, operational test data for the two target location radars were collected over six ranges to the various targets and against four threat levels for the detected and the missed locations; this gave a 2×4×6×2 contingency table for analysis. The interested reader may consult Ref. 37 for further details.

Finally, and with reference to the analysis of contingency tables in general, the Army analyst could well use both the classical chi-square approach and the information theory approach in many of his applications to compare the two to determine which continuing method of analysis is preferable. A new reference to study is that of Gokhale and Kullback (Ref. 38).

5-9 LOGLINEAR ANALYSES OF CONTINGENCY TABLES

In the interest of a more complete and up-to-date account of some of the basic principles for the analysis of contingency tables, many of the models or equations used to determine statistical significance invariably

involve products. Thus it seems quite reasonable to take logarithms and perform the analysis on a loglinear scale. In fact, this would often amount to transforming the original data to a scale that could be more amenable to meeting the assumptions of the analysis of variance technique. Moreover, the Kullback information theory approach to the analysis of contingency tables more or less naturally involves the use of logarithms in a rather basic way, as indicated, for example, in Table 5-11. Indeed, we have already referred to a paper (Ref. 8) that describes and relates the loglinear methods, or models, of analysis to some of the other approaches. Finally, it is also true that the loglinear techniques invariably lead to the ultimate use of equivalent, approximate chi-square values for the determination of statistical significance! Have we not really seen all along in this chapter that although there are what appear to be some different approaches to the problems of the analysis of contingency tables, we appear to wind up with equivalent analyses, more or less? Thus our approach has been through the application of some of the more classical techniques that have been published.

Nevertheless, the loglinear approach does represent a very important recent treatment of contingency table analyses, and many readers will no doubt find wide use of the techniques. In this connection, Fienberg has published a book (Ref. 7) on loglinear methods we heartily recommend to the reader.

As we have indicated, our purpose in this chapter cannot be to discuss extensively each and every method of analysis that the various authors have advanced. In fact, we find especially for contingency tables analyses that one very competent statistician will favor the classical approach, another one may favor the information theory approach, and still another the loglinear approach. Moreover, often there will be very little advantage of one method over the other. Thus we believe and take the position that what may well be needed is a very solid comparison of each approach, perhaps with real data, to show the advantages of one over the other in both the more simple and the multidimensional areas. Nevertheless, we might conclude the loglinear discussion with a useful significance test involving Fisher's odds ratio or the observed "cross-product" ratio.

Fisher's odds ratio, based on the true, underlying p_1 and p_2 for the 2×2 contingency table, was defined in Eq. 5-24. Let us now, for the sake of somewhat shortened notation, define x_{ij} as the observed frequency in the ith row and jth column of a contingency table. (Here $i, j = 1,2$ only.) The observed odds ratio then becomes

$$\hat{\alpha} = x_{11}x_{22}/(x_{12}x_{21}). \tag{5-41}$$

We recall in this connection that if the true $\alpha = 1$, the variables corresponding to rows and columns are independent, whereas if $\alpha \neq 1$, they are dependent or associated. Consequently, we have available a chi-square test for the null hypothesis (Ref. 7), which is

$$\begin{aligned} \chi^2 &= (\ln\hat{\alpha})^2/s^2_{\hat{\alpha}} \\ &= (\ln x_{11}+\ln x_{22}-\ln x_{12}-\ln x_{21})^2 \\ &\quad \times (1/x_{11}+1/x_{12}+1/x_{21}+1/x_{22})^{-1}. \end{aligned} \tag{5-42}$$

Hence it is seen that in terms of a loglinear-type model we also have a very useful approximate chi-square statistic for applications.

The reader is encouraged to read widely on these issues concerning contingency table analyses and to develop the better methods for his own applications.

5-10 SUMMARY

We have presented a discussion of both the classical chi-square approach and the more recent information theory approach by Kullback on the analysis of contingency tables. The very important and widely used 2×2 contingency tables have been covered in some depth to indicate modern methods of analysis, and the concepts of the Fisher exact test, the comparative binomial trials, and the double dichotomy methods of analysis are presented for the analyst. Moreover, for the 2×2 tables, methods of finding confidence bounds on the difference of two proportions, the ratio of the two proportions, and the odds ratio are covered along with tables to apply to these statistical problems.

Analyses of two-way contingency tables by using both the classical chi-square and the Kullback information theory approach are given in sufficient detail so that the Army analyst may have a readily available authentic account for his applications. Finally, some discussion is given of three-way and higher order tables so that the analyst may be prepared to select appropriate literature as required for his particular problems.

Many illustrative examples are presented to give a view of just how the general theory may apply to the analysis of contingency tables in broad Army use.

REFERENCES

1. AMCP 706-111, Engineering Design Handbook, *Experimental Statistics, Section 2, Analysis of Enumerative and Classificatory Data*, December 1969.
2. AMCP 706-114, Engineering Design Handbook, *Experimental Statistics, Section 5, Tables*, December 1969.
3. DARCOM-P 706-101, Engineering Design Handbook, *Army Weapon Systems Analysis, Part One*, November 1977.
4. E. S. Pearson and H. O. Hartley, Eds., *Biometrika Tables for Statisticians* **1**, Cambridge University Press, London, England; published for the Biometrika Trustees, 1954.
5. G. A. Barnard, "Significance Tests for 2×2 Tables", Biometrika **34**, Parts I & II, 123-38 (January 1947).
6. E. S. Pearson, "The Choice of Statistical Tests Illustrated on the Interpretation of Data Classes in a 2×2 Table", Biometrika **34**, Parts I & II, 139-67 (January 1947).
7. Stephen E. Fienberg, *The Analysis of Cross-Classified Categorical Data*, The MIT Press, Cambridge, MA, 1977.
8. H. H. Ku and S. Kullback, "Loglinear Models in Contingency Table Analysis", The American Statistician **28**, 115-22 (November 1974).
9. Herbert Robbins, "A Fundamental Question of Practical Statistics", Letter to the Editor of The American Statistician **31**, 97 (May 1977).
10. Keith R. Eberhardt and Michael A. Fligner, "A Comparison of Two Tests for Equality of Two Proportions", The American Statistician **31**, 151-5 (November 1977).
11. Leo Goodman, "Simultaneous Confidence Intervals for Contrasts Among Multinomial Populations", Annals of Mathematical Statistics **35**, 716-25 (June 1964).
12. E. L. Lehmann, *Testing Statistical Hypotheses*, John Wiley and Sons, New York, NY, 1970.
13. W. J. Conover, Letter to the Editor of The American Statistician **32**, 114 (August 1978).
14. Donald G. Thomas and John J. Gart, "A Table of Exact Confidence Limits for Differences and Ratios of Two Proportions and Their Odds Ratios", Journal of the American Statistical Association **72**, 73-6 (March 1977). (There is a Corrigenda to the Thomas and Gart paper, which appears in Journal of the American Statistical Association **73**, 233 (March 1978).)
15. Thomas J. Santner and Mark K. Snell, "Small-Sample Confidence Intervals for p_1-p_2 and p_1/p_2 in 2×2 Contingency Tables", Journal of the American Statistical Association **75**, 386-94 (June 1980).
16. Mitchell Gail and John J. Gart, "The Determination of Sample Sizes for Use With the Exact Conditional Test in 2×2 Comparative Trials", Biometrics **29**, 441-8 (September 1973).
17. G. R. Garside, "An Accurate Correction for the χ^2 Test in the Homogeneity Case of 2×2 Contingency Tables", New Journal of Statistics and Operations Research **7**, 1-26 (1971).
18. R. D. Boschloo, "Raised Conditional Level of Significance for the 2×2 Table When Testing for the Equality of Two Probabilities", Statistica Neerlandica **21**, 1-35 (1970).
19. G. R. Garside and C. Mack, "Actual Type I Error Probabilities for Various Tests in the Homogeneity Case of the 2×2 Contingency Table", The American Statistician **30**, 18-20 (1976).
20. K. D. Tocher, "Extension of the Neyman-Pearson Theory of Tests to Discontinuous Variates", Biometrika **37**, 130-44 (1950).
21. L. L. McDonald, B. M. Davis, and G. A. Milliken, "A Nonrandomized Unconditional Test for Comparing Two Proportions in 2×2 Contingency Tables", Technometrics **19**, 145-57 (May 1977).

22. M. S. Bartlett, "Contingency Table Interactions", Journal of the Royal Statistical Society, Supplement 2, 248-52 (1935).
23. Solomon Kullback, *The Information in Contingency Tables*, US Army Research Office, Research Triangle Institute, Research Triangle Park, NC, Grant No. DAHCO 4-74-G-0164, September 1974.
24. Solomon Kullback, *Information Theory and Statistics*, John Wiley and Sons, Inc., New York, NY, 1959 (Dover Publications, Inc., New York, NY, 1968 also.).
25. H. H. Ku, R. Varner, and S. Kullback, "Analysis of Multidimensional Contingency Tables", Proceedings of the Fourteenth Conference on the Design of Experiments in Army Research, Development and Testing, US Army Research Office, 141-80, (October 1968).
26. H. H. Ku, R. N. Varner, and S. Kullback, "On the Analysis of Multidimensional Contingency Tables", Journal of the American Statistical Association **66**, 55-64, (March 1971).
27. Solomon Kullback, *Minimum Discrimination Information Estimation and Application*, ARO Report 71-3, Proceedings of the Sixteenth Conference on the Design of Experiments in Army Research, Development and Testing, Sponsored by the Army Mathematics Steering Committee and the US Army Research Office, Research Triangle Park, NC, 1-38, (August 1971).
28. S. Kullback, M. Kupperman, and H. H. Ku, "An Application of Information Theory to the Analysis of Contingency Tables, With a Table of $n\ln n - n$ = 1(1) 10,000", Journal of Research of the National Bureau of Standards—B. Mathematics and Mathematical Physics **66B**, (October-December 1962).
29. Marvin A. Kastenbaum, *Contingency Tables: A Review*, Technical Summary Report No. 596, US Army Mathematics Research Center, Madison, WI, (September 1965).
30. Donald A. Berry and Milton Sobel, "An Improved Procedure for Selecting the Better of Two Bernoulli Populations", Journal of the American Statistical Association **68**, 979-84 (December 1973).
31. J. N. Darroch, "Multiplicative and Additive Interaction in Contingency Tables", Biometrika **61**, 207-14 (1974).
32. Nathan Mantel and Benjamin Hankey, "The Odds Ratio of a 2×2 Contingency Table", The American Statistician **29**, 143-5 (November 1975).
33. John J. Gart, "Point and Interval Estimation of the Common Odds Ratio in the Combination of 2×2 Tables With Fixed Marginals", Biometrika **57**, 471-5 (December 1970).
34. J. B. Copas, "Randomization Models for the Matched and Unmatched 2×2 Tables", Biometrika **60**, 467-76 (1973).
35. J. T. Casagrande, M. C. Pike, and P. G. Smith, "The Power Function of the 'Exact' Test for Comparing Two Binomial Distributions", Applied Statistics **27**, 176-80 (1978).
36. R. Bell, R. E. Mioduski, and E. F. Belbot, *Statistical Analyses of an In-Process Review Questionnaire*, AMSAA Technical Memorandum No. 93, March 1971.
37. Langhorne P. Withers, *Applications of the Fundamental Theorem of Multidimensional Contingency Table Analysis to Operational Test Data*, Proceedings of the Fifteenth Annual US Army Operations Research Symposium, AORS XV, Fort Lee, VA, 26-29 October 1976.
38. B. V. Gokhale and S. Kullback, *The Information in Contingency Tables*, Marcel Dekker, Inc., New York, NY, 1978.
39. Yvonne M. M. Bishop, Stephen E. Fienberg, and Paul W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press, Cambridge, MA, 1975.

BIBLIOGRAPHY

(This Bibliography consists of two parts. The first contains some recent papers that have relevance to some Army contingency table analyses, and the second is a chronological list of references on the general subject of contingency table analyses taken from Kastenbaum (Ref. 29) for the calendar years from 1965 back to 1900.)

Charles C. Brown and Larry R. Muanz, "Reduced Mean Square Error Estimation in Contingency Tables", Journal of the American Statistical Association **71**, 176-82 (March 1976).

Margaret E. Drolete, "The Vanishing 2×2 Table: Linking the Hypergeometric, Binomial and Poisson", The American Statistician **28**, 102-3 (August 1974).

Keith R. Eberhardt and Michael A. Fligner, "A Comparison of Two Tests for Equality of Two Proportions", The American Statistician **31**, 151-5 (November 1977).

Mitchell Gail and Nathan Mantel, "Counting the Number of $r \times c$ Contingency Tables", Journal of the American Statistical Association **72**, 859-62 (December 1977).

Leo A. Goodman, "Guided and Unguided Methods for the Selection of Models for a Set of T Multidimensional Contingency Tables", Journal of the American Statistical Association **68**, 165-75 (March 1973).

Leo A. Goodman, "On the Relation Between Two Statistics Pertaining to Tests of Three-Factor Interaction in Contingency Tables", Journal of the American Statistical Association **70**, 624-5 (September 1975).

William C. Guenther, "Power and Sample Size for Approximate Chi-Square Tests", The American Statistician **31**, 83-5 (May 1977).

M. Halperin, J. H. Ware, D. P. Byar, N. Mantel, C. C. Brown, J. Koziol, M. Gail, and S. B. Green, "Testing for Interaction in an $I \times J \times K$ Contingency Table", Biometrika **64**, 271-5 (August 1977).

A. I. Khuri, "The Shrinking $2 \times c$ Table", (Letter to the Editor) The American Statistician **29**, 68 (February 1975).

S. Keith Lee, "An Example for Teaching Some Basic Concepts in Multidimensional Contingency Table Analysis", The American Statistician **32**, 69-71 (May 1978).

Meena Pradhan and Y. S. Sathe, "Analytical Remarks on Canner's Minimax Method for Finding the Better of Two Binomial Populations", Journal of the American Statistician **71**, 239-41 (March 1976).

Ray E. Schafer, "Modified Estimates for the Binomial Parameter", The American Statistician **30**, 98-100 (May 1976).

Ronald D. Snee, "Graphical Display of Two-Way Contingency Tables", The American Statistician **28**, 9-12 (February 1974).

D. A. Williams, "Improved Likelihood Ratio Tests for Complete Contingency Tables", Biometrika **63**, 33-7 (April 1976).

O. Dale Williams and James E. Grizzle, "Analysis of Contingency Tables Having Ordered Response Categories", Journal of the American Statistical Association **67**, 55-63, (March 1972).

(This part of the Bibliography is taken from Kastenbaum, Ref. 29.)

1965

C. Asano, "On Estimating Multinomial Probabilities by Pooling Incomplete Samples", Annals of Institute of Statistical Mathematics **17**.

I. J. Good, *The Estimation of Probabilities; An Essay on Modern Bayesian Methods*, MIT Research Monograph No. 30, Cambridge, MA.

S. K. Katti and A. N. Sastry, "Biological Examples of Small Expected Frequencies and the Chi-Square Test", Biometrics **21**, 49-54.

R. C. Lewontin and J. Felsenstein, "The Robustness of Homogeneity Tests in $2 \times n$ Tables", Biometrics **21**, 19-33.

V. L. Mote and R. L. Anderson, "An Investigation of the Effect of Misclassification on the Properties of χ^2 Tests in the Analysis of Categorical Data", Biometrika **52**, 95-109.

S. Radhakrishna, "Combination of Results from Several 2×2 Contingency Tables", Biometrics **21**, 86-98.

1964

Harry E. Allison, "Computational Forms for Chi-Square", American Statistician **18**, 17-8.

B. M. Bennett and E. Nakamura, "Tables for Testing Significance in a 2×3 Contingency Table", Technometrics **6**, 439-58.

I. D. J. Bross, "Taking a Covariable Into Account", Journal of the American Statistical Association **59**, 725-36.

Victor Chew, "Application of the Negative Binomial Distribution With Probability of Misclassification", Virginia Journal of Science **15**, 34-40.

Leo A. Goodman, "Simultaneous Confidence Limits for Cross-Product Ratios in Contingency Tables", Journal of the Royal Statistical Society, Series B **26**, 86-102.

Leo A. Goodman, "Simple Methods for Analyzing Three-Factor Interaction in Contingency Tables", Journal of the American Statistical Association **59**, 319-52.

Leo A. Goodman, "Interactions in Multidimensional Contingency Tables", Annals of Mathematical Statistics **35**, 632-46.

Leo A. Goodman, "Simultaneous Confidence Intervals for Contrasts Among Multinomial Populations", Annals of Mathematical Statistics **35**, 716-25.

W. L. Harkness and L. Katz, "Comparison of the Power Functions for the Test of Independence in 2×2 Contingency Tables", Annals of Mathematical Statistics **35**, 1115-27.

D. V. Lindley, "The Bayesian Analysis of Contingency Tables", Annals of Mathematical Statistics **35**, 1622-43.

R. L. Plackett, "The Continuity Correction in 2×2 Tables", Biometrika **21**, Parts 3 and 4, 327-38.

Joseph Putter, "The χ^2 Goodness-of-Fit Test for a Class of Cases of Dependent Observations", Biometrika **51**, 250-2.

R. H. Somers, "Simple Measures of Association for the Triple Dichotomy", Journal of the Royal Statistical Society, Series A **127**, 409-15.

G. M. Tallis, "The Use of Models in the Analysis of Some Classes of Contingency Tables", Biometrics **24**, 832-9.

1963

B. M. Bennett and E. Nakamura, "Tables for Testing Significance in a 2×3 Contingency Table", Technometrics **5**, 501-11.

M. W. Birch, "Maximum Likelihood in Three-Way Contingency Tables", Journal of the Royal Statistical Society, Series B **25**, 220-33.

J. N. Darroch, and S. D. Silvey, "On Testing More Than One Hypothesis", Annals of Mathematical Statistics **34**, 555-67.

Earl L. Diamond, "The Limiting Power of Categorical Data Chi-Square Tests Analogous to Normal Analysis of Variance", Annals of Mathematical Statistics **34**, 1432-41.

A. W. F. Edwards, "The Measure of Association in a 2×2 Table", Journal of the Royal Statistical Society, Series A **126**, 109-14.

S. E. Feldman, and E. Klinger, "Shortcut Calculation of the Fisher-Yates Exact Test", Psychometrika **28**, 289-91.

Ruth A. Gold, "Tests Auxiliary to χ^2 Tests in a Markov Chain", Annals of Mathematical Statistics **34**, 56-74.

I. J. Good, "Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables", Annals of Mathematical Statistics **34**, 911-34.

Leo A. Goodman, "On Methods for Comparing Contingency Tables", Journal of the Royal Statistical Society, Series A **126**, 94-108.

Leo A. Goodman, "On Plackett's Test for Contingency Table Interactions", Journal of the Royal Statistical Society, Series B **25**, 179-88.

Leo A. Goodman and W. H. Kruskal, "Measures of Association for Cross-Classification III: Approximate Sampling Theory", Journal of the American Statistical Association **58**, 310-64.

H. H. Ku, "A Note on Contingency Tables Involving Zero Frequencies and the 2I Test", Technometrics **5**, 398-400.

Nathan Mantel, "Chi-Square Tests With One Degree of Freedom: Extensions of the Mantel-Haenszel Procedure", Journal of the American Statistical Association **58**, 690-700.

D. J. Newell, "Misclassification in 2×2 Tables", Biometrics **19**, 187-8.

Masashi Okamato, "Chi-Square Statistic Based on the Pooled Frequencies of Several Observations", Biometrika **50**, 524-8.

P. N. Ries, and Harry Smith, "The Use of Chi-Square for Preference Testing in Multidimensional Problems", Chemical Engineering Program Symposium Series **59**, 39-43.

John E. Walsh, "Loss in Test Efficiency Due to Misclassification for 2×2 Tables", Biometrics **19**, 158-62.

1962

C. Daly, "A Simple Test for Trends in a Contingency Table", Biometrics **18**, 114-9.

J. N. Darroch, "Interactions in Multifactor Contingency Tables," Journal of the Royal Statistical Society, Series B **24**, 251-63.

Sir Ronald A. Fisher, "Confidence Limits for a Cross-Product Ratio", Austrialian Journal of Statistics **4**, 41.

J. J. Gart, "Approximate Confidence Limits for Relative Risks", Journal of the Royal Statistical Society, Series B **24**, 454-63.

J. J. Gart, "On the Combination of Relative Risks", Biometrics **18**, 601-10.

W. M. Kincaid, "The Combination of 2×*m* Contingency Tables", Biometrics **18**, 224-8.

S. Kullback, M. Kupperman, and H. H. Ku, "An Application of Information Theory to the Analysis of Contingency Tables With a Table of $2N\ln N$, $N = 1(1)\ 10{,}000$", Journal of Research National Bureau of Standards B **66**, 217-43.

S. Kullback, M. Kupperman, and H. H. Ku, "Tests for Contingency Tables and Markov Chains", Technometrics **4**, 573-608.

B. N. Lewis, "On the Analysis of Interaction in Multidimensional Contingency Tables", Journal of the Royal Statistical Society, Series A **125**, 88-117.

R. L. Plackett, "A Note on Interactions in Contingency Tables", Journal of the Royal Statistical Society, Series B **24**, 162-6.

G. M. Tallis, "The Maximum Likelihood Estimation of Correlation from Contingency Tables", Biometrics **18**, 342-53.

1961

Agnes Berger, "On Comparing Intensities of Association Between Two Binary Characteristics in Two Different Populations", Journal of the American Statistical Association **56**, 889-908.

V. P. Bhapkar, "Some Tests for Categorical Data", Annals of Mathematical Statistics **32**, 72-83.

P. Billingsley, "Statistical Inference for Markov Processes", Statistical Research Monographs **2**, The University of Chicago Press.

P. J. Claringbold, "The Use of Orthogonal Polynomials in the Partition of Chi-Square", Australian Journal of Statistics **3**, 48-63.

D. Friedlander, "A Technique for Estimating a Contingency Table, Given the Marginal Totals and Some Supplementary Data", Journal of the Royal Statistical Society, Series A **124**, 412-20.

G. Gregory, "Contingency Tables With a Dependent Classification", Australian Journal of Statistics **3**, 42-7.

James E. Grizzle, "A New Method of Testing Hypotheses and Estimating Parameters for the Logistic Model", Biometrics **17**, 372-85.

M. G. Kendall and A. Stuart, The Advanced Theory of Statistics **2**, Charles Griffin and Company, London, England.

M. Okamato and G. Ishii, "Test of Independence in Intraclass 2×2 Tables", Biometrika **48**, 181-90.

E. Rogot, "A Note on Measurement Errors and Detecting Real Differences", Journal of the American Statistical Association **56**, 314-19.

William J. Schull, "Some Problems of Analysis of Multifactor Tables", Bulletin of the International Statistics Institute **28**, Part 3, 259-70.

F. Yates, "Marginal Percentages in Multiway Tables of Quantal Data With Disproportionate Frequencies", Biometrics **17**, 1-9.

1960

B. M. Bennett and P. Hsu, "On the Power Function of the Exact Test for the 2×2 Contingency Table", Biometrika **47**, 393-8.

N. T. Gridgeman, "Card-Matching Experiments: A Conspectus of Theory", Journal of the Royal Statistical Society, Series A **123**, 45-9.

G. Ishii, "Intraclass Contingency Tables", Annals of the Institute of Statistical Mathematics **12**, 161-207; corrections, 279.

M. A. Kastenbaum, "A Note on the Additive Partitioning of Chi-Square in Contingency Tables", Biometrics **16**, 416-22.

Morton Kupperman, "On Comparing Two Observed Frequency Counts", Applied Statistics **9**, 37-42.

H. O. Lancaster, "On Tests on Independence in Several Dimensiona", Journal of the Australian Mathematical Society **1**, 241-54.

W. H. Robertson, "Programming Fisher's Exact Method of Comparing Two Percentages", Technometrics **2**, 103-7.

1959

R. L. Anderson, "Use of Contingency Tables in the Analysis of Consumer Preference Studies", Biometrics **15**, 582-90.

I. M. Chakravarti and C. R. Rao, "Tables for Some Small Sample Tests of Significance for Poisson Distributions and 2×3 Contingency Tables", Sankhya **21**, Parts 3 and 4, 315-26.

Leo A. Goodman and W. H. Kruskal, "Measures of Association for Cross-Classification II: Further Discussion and References", Journal of the American Statistical Association **54**, 123-63.

J. B. S. Haldane, "The Analysis of Heterogeneity, I", Sankhya **21**, Parts 3 and 4, 209-16.

C. J. Hoyt, P. R. Krishnaich, and E. P. Torrance, "Analysis of Complex Contingency Data", Journal of Experimental Education **27**, 187-94.

M. A. Kastenbaum and D. E. Lamphiear, "Calculation of Chi-Square to Test the No Three-Factor Interaction Hypothesis", Biometrics **15**, 107-15.

S. Kullback, *Information Theory and Statistics,* John Wiley and Sons, New York, NY.

Morton Kupperman, "A Rapid Significance Test for Contingency Tables", Biometrics **15**, 625-8.

C. A. G. Nass, "The χ^2 Test for Small Expectations in Contingency Tables, With Special Reference to Accidents and Absenteeism", Biometrika **46**, 365-85.

S. D. Silvey, "The Lagrangian Multiplier Test", Annals of Mathematical Statistics **30**, 389-407.

Robert H. Somers, "The Rank Analogue of Product-Moment Partial Correlation and Regression, With Application to Manifold, Ordered Continency Tables", Biometrika **46**, 241-6.

H. S. Steyn, "On χ^2 Tests for Contingency Tables of Negative Binomial Type", Statistics Neerlandica **13**, 433-44.

Irving B. Weiner, "A Note on the Use of Mood's Likelihood Ratio Test for Item Analyses Involving 2×2 Tables With Small Samples", Psychometrika **24**, 371-2.

1958

H. M. Blalock, Jr., "Probablistic Interpretations for the Mean Square Contingency", Journal of the American Statistical Association **53**, 102-5.

M. A. Kastenbaum, "Estimation of Relative Frequencies of Four Sperm Types in Drosophila Melanogaster", Biometrics **14**, 223-8.

S. K. Mitra, "On the Limiting Power Function of the Frequency Chi-Square Test", Annals of Mathematical Statistics **29**, 1221-33.

G. W. Snedecor, "Chi-Square of Bartlett, Mood and Lancaster in a 2^3 Contingency Table", Biometrics **14**, 560-2 (Query).

1957

Irwin D. J. Bross and Ethel L. Kasten, "Rapid Analysis of 2×2 Tables", Journal of the American Statistical Association **52**, 18-28.

L. C. A. Corsten, "Partition of Experimental Vectors Connected with Multinomial Distributions", Biometrics **13**, 451-84.

J. H. Edwards, "A Note on the Practical Interpretation of 2×2 Tables", British Journal of the Preventive Society of Medicine, **11**, 73-8.

H. O. Lancaster, "Some Properties of the Bivariate Normal Distribution Considered in the Form of a Contingency Table", Biometrika **44**, 289-92.

V. L. Mote, "An Investigation of the Effect of Misclassification of the Chi-Square Tests in the Analysis of Categorical Data", Unpublished PhD dissertation, N.C. State University, Raleigh, NC (Also Institute of Statistics Mimeo Series No. 182).

S. N. Roy, *Some Aspects of Multivariate Analysis*, John Wiley and Sons, Inc., New York, NY.

J. M. Sakoda and B. H. Cohen, "Exact Probabilities for Contingency Tables Using Binomial Coefficients", Psychometrika **22**, 83-6.

Barnet Woolf, "The Log Likelihood Ratio Test (The G-Test). Methods and Tables for Tests of Heterogeneity in Contingency Tables", Annals of Human Genetics **21**, 397-409.

1956

J. A. Fishman, "A Note on Jenkins' Improved Method for Tetrachoric *r*", Psychometrika **20**, 305.

I. J. Good, "On the Estimation of Small Frequencies in Contingency Tables", Journal of the Royal Statistical Society, Series B **13**, 113-24.

N. T. Gridgeman, "A Tasting Experiment", Applied Statistics **15**, 106-12.

E. K. Leander and D. J. Finney, "An Extension of the Use of the χ^2 Test", Applied Statistics **5**, 132-6.

D. Mainland, L. Herrera, and M. I. Sutcliffe, "Statistical Tables for Use With Binomial Samples—Contingency Tests, Confidence Limits, and Sample Size Estimates", New York University College of Medicine, New York, NY.

S. N. Roy and Marvin A. Kastenbaum, "On the Hypothesis of No 'Interaction' in a Multiway Contingency Table", Annals of Mathematical Statistics **27**, 749-57.

S. N. Roy and S. K. Mitra, "An Introduction to Some Nonparametric Generalizations of Analysis of Variance and Multivariate Analysis", Biometrika **43**, Parts 3 and 4, 361-76.

G. S. Watson, "Missing and 'Mixed-Up' Frequencies in Contingency Tables", Biometrics **12**, 47-50.

1955

P. Armitage, "Tests for Linear Trends in Proportions and Frequencies", Biometrics **11**, 375-86.

P. Armsen, "Tables for Significance Tests of 2×2 Contingency Tables", Biometrika **42**, 494-505.

W. G. Cochran, "A Test of a Linear Function of the Deviations Between Observed and Expected Numbers", Journal of the American Statistical Association **50**, 377-97.

J. B. S. Haldane, "Substitutes for χ^2", Biometrika **42**, 265-6.

J. B. S. Haldane, "A Problem in the Significance of Small Numbers", Biometrika **42**, 266-7.

J. B. S. Haldane, "The Rapid Calculation of χ^2 as a Test of Homogeneity from a $2\times n$ Table", Biometrika **42**, 519-20.

W. L. Jenkins, "An Improved Method for Tetrachoric *r*", Psychometrika **20**, 253-8.

Marvin A. Kastenbaum, "Analysis of Data in Multiway Contingency Tables", Unpublished PhD dissertation, North Carolina State University, Raleigh, NC, October 1955.

P. H. Leslie, "A Simple Method of Calculating the Exact Probability in 2×2 Contingency Tables With Small Marginal Totals", Biometrika **42**, 522-3.

S. K. Mitra, "Contributions to the Statistical Analysis of Categorical Data", North Carolina Institute of Statistics Mimeograph Series, No. 142, December 1955.

S. N. Roy and Marvin A. Kastenbaum, "A Generalization of Analysis of Variance and Multivariate Analysis to Data Based on Frequencies in Qualitative Categorical or Class Intervals", North Carolina Institute of Statistics Mimeograph Series, No. 131, June 1955.

S. N. Roy and S. K. Mitra, "An Introduction to Some Nonparametric Generalizations of Analysis of Variance and Multivariate Analysis", North Carolina Institute of Statistics Mimeograph Series, No. 139, November 1955.

C. Chandra Sekar, S. P. Agarivala, and P. N. Chakraborty, "On the Power Function of a Test of Significance for the Difference Between Two Proportions", Sankhya **15**, Part 4, 381-90.

Alan Stuart, "A Test of Homogeneity of the Marginal Distributions in a Two-Way Classification", Biometrika **42**, 412-6.
Barnet Woolf, "On Estimating the Relation Between Blood Group and Disease", Annals of Human Genetics **19**, 251-3.
F. Yates, "A Note on the Application of the Combination of Probabilities Test to a Set of 2×2 Tables", Biometrika **42**, 401-11.
F. Yates, "The Use of Transformations and Maximum Likelihood in the Analysis of Quantal Experiments Involving Two Treatments", Biometrika **42**, 382-403.

1954

Irwin D. J. Bross, "Misclassification in 2×2 Tables", Biometrics **10**, 478-86.
W. G. Cochran, "Some Methods for Strengthening the Common Chi-Square Tests", Biometrics **10**, 417-51.
R. B. Dawson, "A Simplified Expression for the Variance of the χ^2 Function on a Contingency Table", Biometrika **41**, 280.
Leo A. Goodman and W. H. Kruskal, "Measures of Association for Cross-Classification", Journal of the American Statistical Association **49**, 732-64.
A. W. Kimball, "Shortcut Formulas for the Exact Partition of Chi-Square in Contingency Tables", Biometrics **10**, 452-8.
W. J. McGill, "Multivariate Information Transmission", Psychometrika **19**, 97-116.

1952

William G. Cochran, "The χ^2 Test of Goodness of Fit", Annals of Mathematical Statistics **23**, 315-45.
G. V. Dyke and H. D. Patterson, "Analysis of Factorial Arrangements When the Data Are Proportions", Biometrics **8**, 1-12.

1951

H. O. Lancaster, "Complex Contingency Tables Treated by the Partition of Chi-Square", Journal of the Royal Statistical Society, Series B **13**, 242-9.
C. H. Simpson, "The Interpretation of Interaction in Contingency Tables", Journal of the Royal Statistical Society, Series B **13**, 238-41.

1950

K. D. Tocher, "Extension of the Neyman-Pearson Theory of Tests to Discontinuous Variates", Biometrika **37**, 130-44.

1949

P. L. Hsu, "The Limiting Distributions of Functions of Sample Means and Application to Testing Hypotheses", Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley and Los Angeles, CA, 1945-1946.
J. O. Irwin, "A Note on the Subdivision of Chi-Square Into Components", Biometrika **36**, 130-34.
H. O. Lancaster, "The Derivation and Partition of Chi-Square in Certain Discrete Distributions", Biometrika **36**, 117-29.

1948

F. Yates, "The Analysis of Contingency Tables With Groupings Based on Quantitative Characters", Biometrika **35**, 176-81.

1947

G. A. Barnard, "Significance Tests for 2×2 Tables", Biometrika **34**, 123-38.
E. S. Pearson, "The Choice of Statistical Tests Illustrated on the Interpretation of Data Classed in a 2×2 Table", Biometrika **34**, 139-67.

1946

H. Cramer, *Mathematical Methods of Statistics*, Princeton University Press, 424, Princeton, NJ.

1945

H. W. Norton, "Calculation of Chi-Square for Complex Contingency Tables", Journal of the American Statistical Association **40**, 251-8.

1937

J. B. S. Haldane, "The Exact Value of the Moments of the Distribution of χ^2 Used as a Test of Goodness of Fit, When Expectations Are Small", Biometrika **29**, 133-43.

1935

M. S. Bartlett, "Contingency Table Interactions", Journal of the Royal Statistical Society Supplement 2, 248-52.

S. S. Wilks, "The Likelihood Test of Independence in Contingency Tables", Annals of Mathematical Statistics **6**, 190-6.

1934

R. A. Fisher, *Statistical Methods for Research Workers*, Edinburgh 5th and subsequent editions, Oliver and Boyd, Ltd., Section 21.02, London, England.

F. Yates, "Contingency Tables Involving Small Numbers and the χ^2 Test", Journal of the Royal Statistical Society Supplement 1, **1**, 217-35.

1924

R. A. Fisher, "The Conditions Under Which Chi-Square Measures the Discrepancy Between Observation and Hypothesis", Journal of the Royal Statistical Society **87**, 442-50.

1922

R. A. Fisher, "On the Interpretation of Chi-Square from Contingency Tables, and the Calculation of *P*", Journal of the Royal Statistical Society **85**, 87-94.

1900

Karl Pearson, "On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling", Philosophy Magazine, Series 5 **50**, 157-72.

CHAPTER 6

LEAST SQUARES, REGRESSION, AND FUNCTIONAL RELATIONS

The use of least squares procedures to fit lines, curves, or functional relations to observational data represents one of the oldest forms of statistical endeavor. The topics of least squares and regression, therefore, are presented in considerable statistical detail so that the analyst can have available a very comprehensive coverage of the subject. Presented first are the simpler concepts of fitting a line to data for the case in which only the dependent variable is subject to error and the independent variable is free of error. This is extended to cover the case in which for linear fits both the dependent and independent variables are subject to errors of determination. Estimation problems and the use of appropriate significance tests are discussed in detail.

The fitting of planes, hyperplanes, and polynomials are next covered in detail, along with the case in which the independent variables are spaced equally and orthogonal polynomials can be used.

Functional relationships, or the physics of the applications, are stressed along with least squares procedures in order to obtain sound predictive equations. Moreover, in nonlinear or generalized least squares problems relating to particular applications, the clever choice of the function form may lead to best results, especially for the important practical case of errors in both dependent and independent variables.

Many applications of the theory to observed data are discussed in the form of examples.

6-0 LIST OF SYMBOLS

A_i = transformed coefficients determined for the orthogonal polynomials $P_r(t_i)$ as in Eq. 6-123

$A_{xx} = n\Sigma x^2 - (\Sigma x)^2$

$A_{xy} = n\Sigma xy - (\Sigma x)(\Sigma y)$ (may use any letter subscripts)

a = estimate of α

a_i = coefficient of polynomial term in Eq. 6-132

a_0 = value of a

BHN = Brinell hardness number

BL = ballistic limit of armor

BL = barrel length, in.

b = estimate of β, as is $\hat{\beta}$

b_i = original coefficients in a polynomial, $i = 0,1,2, \ldots$, as in Eq. 6-121

b_j = estimate of the β_j

b_{x_y} = slope of regression line of x on y

b_{y_x} = slope of regression line of y on x

$[C]$ = inverse matrix given in Eq. 6-147

Cov = denotes covariance of a quantity

c = constant

c = estimate of γ

c_{ij} = represents the ijth element of the inverse matrix $[C]$

$[D_i]$ = denotes the ith iterative stage of a computation to determine the vector value $[\mu]$—see Eqs. 6-170 through 6-173

d = constant

d_i = ith error in y, i.e., for the observation $y_i = \eta_i + d_i$

$dx_i = x_i - x_{i-1}$ = first forward difference in the x_i

$dy_i = y_i - y_{i-1}$ = first forward difference in the y_i

$E(\ )$ = expected value of ( )
e = constant value
e = designation of error
e = q vector of errors, some of which could be zero
e_i = error (of measurement) in x_i, if applicable
$F_0(2,n-2)$ = function in Eq. 6-27 following the F distribution
F_γ = $F_\gamma(2,n-2)$ = upper γ probability level of the Fisher-Snedecor F distribution for 2 and $(n-2)$ degrees of freedom
f = $f(z,\theta)$ = k vector of functional forms
f = constant value (or a function)
$[f]$ = $[f(x,\mu)]$ = vector of functions
$f(\mu)$ = function of the true part of the independent variable x when it contains error; the function f is fitted to data.
f_z = $f_z(z,\theta)$ = Jacobian matrix of partial derivatives of f with respect to z
f_θ = $f_\theta(z,\theta)$ = Jacobian matrix of partial derivatives of f with respect to θ
$[f'(\mu)]$ = denotes the Jacobian matrix of Eq. 6-169
h = constant
I_1 = first instrument
I_2 = second instrument
k = constant (also degree of a polynomial)
k = denotes the number of components of a functional vector (k is a scalar)
M_R = mass of residual fragment or projectile
MV = muzzle velocity
m_s = striking mass of projectile against armor
$N(0,\sigma^2)$ = designates a normal distribution with mean of zero and variance σ^2
n = sample size
$P_r(t_i)$ = orthogonal polynomial in t_i for the rth degree (likewise for s in place of r) $r = 0,1,2,...$
p = denotes the number of parameters fitted in least squares (p is a scalar)
plim = probability limit
q = denotes the number of components of the z vector (q is a scalar)
R = variance-covariance matrix of the errors e
r = degree of a polynomial
r = r_{xy} = sample correlation coefficient
r = designates readings of the first instrument I_1
S_{dxdy} = sample covariance of dx's and dy's
S_{xy} = sample covariance of x and y = $A_{xy}/[n(n-1)]$
S^2 = $S^2_{y_x}$ = sample variance of residuals from least squares fit, i.e., observed minus fitted points
S^2_{dx} = sample variance of the differences dx_i
S^2_{dy} = sample variance of the differences dy_i
S^2_x = sample variance of x (likewise for other subscripts)
s = designates readings of the second instrument I_2 (two or more instruments)
t_b = Student's t for the subscript b—similar for a or other letter
t_i = linear transformation of the x_i for orthogonal polynomials
$t_{\gamma/2}(n-2)$ = upper $\gamma/2$ probability level of Student's t for $(n-2)$ degrees of freedom
u_i = independent variable

$\text{Var}(b) = \sigma_b^2 = E(b - \beta)^2 =$ variance of b

$V_R =$ residual velocity

$V_S =$ striking velocity

$\nu_i =$ independent variable

$[X] =$ used to denote n observational values of the independent variable x in polynomial form as in Eq. 6-151

$[X]_0^T =$ general type of vector representing either the linear form of the x_i in Eq. 6-149 or components of an $(r - 1)$st power polynomial in x as in Eq. 6-150. No observations on x are included.

$x =$ usually an independent variable

$x^* =$ preselected or standard value of x

$x_{ij} = i$th measurement of the jth independent variable

$x_0 =$ specified value of x

$\bar{x} = \Sigma x/n =$ mean of x

$(\bar{x}_1, \bar{y}_1) =$ coordinates of the means of the lower third of n pairs of points (x_i, y_i) for $i=1, 2,..., n$

$(\bar{x}_3, \bar{y}_3) =$ coordinates of the means of the upper third of the points (x, y)

$\bar{x}^2 = \Sigma x^2/n =$ mean value of the x^2 observations

$y =$ usually a dependent variable

$y' =$ another value of or designation for y

$y_i = i$th (dependent variable) observation

$z =$ letter to denote a dependent variable when x and y are independent variables

$z_i = i$th iterative stage for the vector z

$z_m =$ vector of measurements on the dependent and independent variables

$z_t =$ true values of z when z is subject to error and is used as a q vector

$\alpha =$ constant intercept true value

$\beta =$ true slope of a line

$\beta_j =$ true unknown coefficients of the linear regression terms

$\beta_0 =$ specified value of β

$\hat{\beta} =$ an estimate of β

$\gamma =$ true unknown coefficient or a probability level

$\Delta_1 =$ determinant of A_{xy}-type calculations

$\delta =$ true unknown coefficient

$\eta_i =$ true unknown part or component of y_i

$\eta_0 =$ specified value of η

$\theta = p$ vector of unknown parameters in generalized least squares

$\theta_i = i$th iterative stage for the vector θ

$\lambda = \sigma_d^2/\sigma_e^2 =$ ratio of variances of errors in y to errors in x

$\lambda_i =$ coefficients used in orthogonal polynomials of Eq. 6-130

$[\mu] =$ denotes a vector of components μ_i of μ

$\mu_i =$ true value of the independent variable for the ith observation—free of error

$\mu_i =$ denotes the ith iterative stage of $[\mu]$

$\xi_i' = \lambda_i \xi_i =$ transformations of the t_i as in Eq. 6-130

$\rho =$ population correlation coefficient

$\rho_1 =$ designates the population serial correlation coefficient of lag 1

$\sigma_{bc} =$ population covariance of the estimates b and c of β and γ, respectively

σ_{ey} = standard error of measurement in y (the first subscript means "error")
σ_{xy} = true covariance of x and y
$\hat{\sigma}$ = estimate of σ
σ^2 = population variance
$\sigma^2_{y_x} = \sigma^2_d = \sigma^2$ = population variance of errors d_i or residuals
ϕ = function of observed minus fitted values of the sum of squares to be minimized in Eq. 6-5
$\partial\phi/\partial a$ = derivative of ϕ with respect to a (b may be substituted for a)
$[\ ]$ = denotes a vector or matrix
$[\ ]^T$ = denotes the transpose of a vector or matrix. (The transpose of a column vector gives a row vector.)
$[\ ]^{-1}$ = denotes the inverse of a matrix

6-1 INTRODUCTION

A frequent and important practical problem in research and development is to determine an appropriate relationship, or the best fitting law, between variables of interest, i.e., fitting equations to data, and testing various hypotheses concerning the physical values or the relation of the parameters studied. In addition, and as usual, we would like to summarize experimental data in the form of an equation or "law" and be able to predict future or expected occurrences from our fitted or empirically determined law. Indeed, in many problems it is important to be able to place confidence bounds on the various physical parameters that can be estimated or inferred from the data developed in an experiment.

Needless to say, this is a more involved problem than it may appear initially. Indeed, one should expect that errors of measurement will be made in practically all determinations of the values of the variables in any experiment. Also in many cases we encounter the additional problem of properly treating the random or unaccounted-for variations in addition to the underlying physical laws—or functional relations—we seek to sort out of the "noise". Of course, we might say that we would prefer to establish a law of enduring relationship between the key variables or parameters of interest, which is actually free of any measurement error or other variations of extraneous interest. In addition, it becomes important to know just how precise or accurate our final prediction is since it might be desirable to conduct more experiments, but this would depend especially on our subsequent uses of the fitted equation. A general but simple and enduring law makes a very definite contribution to science and technology.

We should remark initially and keep in mind that the practice of transforming variables to linear functions or relations, as is often done in the physical sciences or in engineering—i.e., attempts toward "linearizing the data"—is an excellent one indeed, as we will see in the sequel, because it helps to establish relationships between complex quantities and to simplify much of the resulting analysis. Furthermore, it usually is not difficult to transfer statistical or physical statements about the transformed data back to equivalent ones about the original variables. For this reason, we will cover the case of linear least squares, or linear regression, in considerable detail and then consider the functional or "structural" relations of the variables involved. We will, therefore, start with the case of the simple linear regression between an independent variable that is assumed to be free of measurement error and the dependent variable that is measured or found with error of determination. After covering some particular points of practical significance, we will proceed to a discussion of the more complex cases. It is highly desirable in this connection to distinguish between "controlled" or "fixed" variables, random variables or variates, and the errors of measurement that may be either of a random or systematic nature.

Chapter 5 (Ref. 1) contains an excellent introduction and very useful account of the problem of fitting straight lines to data. In fact, it gives step-by-step procedures that may be easily followed along with all of the statistical tests of significance needed for a rather complete linear analysis. Hence in our approach we will repeat only that coverage of Ref. 1 deemed necessary to review or to establish a sufficient background for more advanced topics needed to update the contents of Ref. 1 for more recent applications. Also we will discuss some especially useful aspects of regression and curve fitting not included in Ref. 1 and will emphasize the more modern statistical analyses of possible errors of measurement in one or both variables. Moreover, we

will dwell at some length on linear least squares since they will continue to be very widely applied and the linear methods are prerequisite to the analysis of many of the nonlinear techniques.

In many ways our approach to least squares and curve fitting is different from the usual methods or forms of computation practiced as a result of some of the usual textbooks on statistics. We recommend a rather special form of key parameters in the course of the calculations that are free of rounding error until the last few steps. This, we believe, is an advantage in many applications.

6-2 LINEAR LEAST SQUARES OR REGRESSION FOR A DEPENDENT VARIABLE (MEASURED WITH ERROR) AND AN INDEPENDENT VARIABLE (WITHOUT ERROR)

6-2.1 GENERAL

In dealing with experimental data involving two variables x and y—for example, time and distance measurements or muzzle velocity and range relations—there may appear to be a trend or some mathematical relation (linear or otherwise) between the plotted values of x and y. We will therefore be interested in estimating the best relation between x and y and in judging statistically whether or not the determined relation is a significant one. The method used is generally referred to as the "least squares" technique, i.e., the process of finding an appropriate "regression" of y on x, although there are other methods of fitting a selected law between two or more variables, e.g., the technique of maximum likelihood (ML). In the method of least squares, we assume a model or relation between the variables—such as the linear, quadratic, or exponential forms—which involves certain unknown parameters or coefficients, and then fit the hypothesized curve to the two or more variables so that the sum of squares (SS) of the residuals or (vertical) deviations from the fitted curve is a minimum. The significance of the fitted curve, or its key parameters, will be tested statistically and otherwise established. Also if considered desirable, confidence bounds may be placed on the estimated parameters or coefficients, the fitted curve, and the predictions for future observations.

Our approach will consist of combining the physical and statistical points of view insofar as possible. Thus our models and assumptions will consider both the functional or structural relation between true values of the variables and the statistical treatment of variates or errors of measurement and their probability distributions. In the model of this paragraph the independent variable is assumed to be free of error, and hence only the dependent y variable is subject to error.

6-2.2 THE LINE—ONE VARIABLE (y) SUBJECT TO ERROR

Suppose we are dealing with two observable variables, x and y, which are connected by an apparent linear relation. Suppose further that the dependent variable y not only depends on x but is also subject to (random) errors of measurement. That is, y as measured physically includes an error of measurement, whereas x is a controlled or "fixed" (mathematical) variable that is free of any measurement errors or almost completely free of errors as compared to the measured dependent variable y. Over the interval of physical interest in an experiment, it will be assumed that the variability, or the variance, in the errors of y is essentially constant. The mean value of y depends on the value of x considered, and the variance of y about the hypothesized linear relation is independent of the value of x, i.e., the variance or standard error about the hypothesized linear relation or fitted line is independent of the value of x, i.e., constant over the range of x used in the experiment.

To illustrate some of these points more clearly, we have selected a particular, yet rather simple, example from the American Society for Testing and Materials (ASTM) *Manual on Fitting Straight Lines* (Ref. 2). The observed data were obtained in a calibration experiment of a new method (gravimetric determination) for estimating the amount of calcium in the presence of large amounts of magnesium. The experimental data are given in Table 6-1 for known amounts of CaO (x) and the observed amounts of CaO found by the new method (y). Thus we can say that x is free of (measurement) error and that the new method y may be subject to errors of determination.

The basic reasons for selecting this particular example should be clear—the independent variable x should be quite free of error and the dependent variable y for any new method should be judged along with the known x in order to study its properties, especially to learn of its precision and accuracy in case the new method is

TABLE 6-1
GRAVIMETRIC DETERMINATION OF CALCIUM IN THE PRESENCE OF MAGNESIUM

x CaO Actually Present, mg	y CaO Found by New Method, mg
20.0	19.8
22.5	22.8
25.0	24.5
28.5	27.3
31.0	31.0
33.5	35.0
35.5	35.1
37.0	37.1
38.0	38.5
40.0	39.0

Copyright, ASTM, 1916 Race Street, Philadelphia, PA 19103. Reprinted/Adapted with permission.

adopted. A plot of y against x would indicate a nearly linear relation, as it should. Also since x and y may be considered to be measurements of the same quantity, the slope of the fitted line should be 45 deg, and moreover, the line should pass through the origin for the assumption of linearity and good calibration of both methods. In addition, the error of determination or measurement of the new method should be acceptable. It is our purpose, therefore, to consider each of these questions in detail.

Furthermore, we should remark that the measured x and y are not random variables, but there is a physical (linear) or mathematical relation between the two. In this particular calibration experiment, the CaO actually present, or x, has been varied purposely over the range so that y will correspondingly vary but with the probable addition of random measurement errors. In fact, the precision of measurement of the new method y could be determined by the techniques of Chapter 2 because those models include the measurements of the same quantities. However, we will delay any such calculations using the methods of Chapter 2 until we have fitted the line.

The n observed values of x and y are represented algebraically by (x_1, y_1), (x_2, y_2), (x_3, y_3),..., (x_i, y_i),..., (x_n, y_n).

The linear model or assumption considered for the relation between x and y, i.e., the observed pairs (x_i, y_i), is

$$x_i = \mu_i \tag{6-1}$$

$$y_i = \alpha + \beta\mu_i + d_i = \eta_i + d_i \tag{6-2}$$

where

μ_i = true value of the independent variable for the ith observation—free of error
α = constant intercept true value
β = true slope of line
d_i = ith error in y, i.e., for the observation $y_i = \eta_i + d_i$
η_i = true unknown part, or component, of y_i.

We use the notation of Eq. 6-2 to indicate that the measured value y contains a true part η_i and possibly an error of measurement designated by d_i. Moreover, x_i is considered to be free of any measurement error since we can set its true value μ_i in this case. (If x_i were to contain an error of measurement under the hypothesis, we would write it as $x_i = \mu_i + e_i$, in which the first factor is the true value and the second is an error in the measurement of x.) The relation given by

$$\eta = \alpha + \beta\mu \tag{6-3}$$

is called the true (functional) relation between the parts of x and y in which we are interested. It is also the true regression in our simple model.

The errors d_i have mean or expected value, $E(d_i) = 0$, and variance in the errors $E[d_i - E(d_i)]^2 = \sigma_d^2 = \sigma_{y_x}^2$ or simply σ^2, the constant variance about the fitted regression line.

Thus the mean value of an observed y for a given value of x is

$$E(y) = E(\alpha + \beta x + d) = \alpha + \beta\bar{x} = \alpha + \beta\mu$$

as in Eq. 6-3.

The variance of y about its population mean, $\alpha + \beta x = \alpha + \beta\mu$, is $E(y - \alpha - \beta x)^2 = E(d_i^2) = \sigma_{y_x}^2 = \sigma_d^2$, i.e., the population "variance of residuals", or the variance of an individual observation about the regression line.

Of course, for a small sample of n observed pairs (x_i, y_i), it will not be possible to estimate α and β very precisely. Our *fitted* line will therefore be of the form

$$y = a + bx \tag{6-4}$$

where a and b are estimates of α and β, respectively, and are therefore subject to "error" or statistical variation.

We estimate α and β from a and b, respectively, by determining a and b so that

$$\phi = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \tag{6-5}$$

is a minimum.

Now

$$\frac{\partial\phi}{\partial a} = -2\sum_{i=1}^{n}(y_i - a - bx_i) = -2(\Sigma y_i - na - b\Sigma x_i) \tag{6-6}$$

and we find also that

$$\frac{\partial\phi}{\partial b} = -2\sum_{i=1}^{n}(y_i - a - bx_i)x_i = -2(\Sigma x_i y_i - a\Sigma x_i - b\Sigma x_i^2). \tag{6-7}$$

Equating $\partial\phi/\partial a$ and $\partial\phi/\partial b$, respectively, to zero, we obtain the well-known *normal* equations:

$$na + (\Sigma x_i)b = \Sigma y_i \tag{6-8}$$

$$(\Sigma x_i)a + (\Sigma x_i^2)b = \Sigma x_i y_i. \tag{6-9}$$

Solving Eqs. 6-8 and 6-9 for a and b, we find

$$a = \text{est}\alpha = \frac{(\Sigma y_i)(\Sigma x_i^2) - (\Sigma x_i y_i)(\Sigma x_i)}{A_{xx}}$$

$$= \bar{y} - b\bar{x}, \text{ or } \frac{1}{n}(\Sigma y_i - b\Sigma x_i) \tag{6-10}$$

$$b = \text{est}\beta = \frac{A_{xy}}{A_{xx}} \tag{6-11}$$

where

$$A_{xx} = n\Sigma x_i^2 - (\Sigma x_i)^2 \tag{6-12}$$

$$A_{xy} = n\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i). \tag{6-13}$$

These quantities are established for computational purposes since they may be used free of rounding error and have advantages the reader will appreciate in what follows.

The variance of residuals $\sigma^2_{y_x} = \sigma^2$, i.e., the variance of an individual deviation from the fitted line, is estimated from

$$S^2_{y_x} = S^2 = \left(\frac{1}{n-2}\right)\sum_{i=1}^{n}(y_i - a - bx_i)^2 = \frac{\Sigma y_i^2 - a\Sigma y_i - b\Sigma x_i y_i}{n-2} = \frac{1}{n(n-2)}\left(A_{yy} - \frac{A_{xy}^2}{A_{xx}}\right). \tag{6-14}$$

The quantity

$$r = r_{xy} = \frac{A_{xy}}{\sqrt{A_{xx}A_{yy}}} \tag{6-15}$$

is called the product moment correlation coefficient. For very large samples

$$\sigma^2_{y_x} = \sigma^2_d = \sigma^2_y(1 - \rho^2)* \tag{6-16}$$

where ρ is the population correlation coefficient between the variables x and y. Note that also

$$\beta = \rho\sigma_y/\sigma_x. \tag{6-17}$$

where

σ_x = standard deviation of x
σ_y = standard deviation of y.

Now it can be shown that the mean, or expected, values of a and b are α and β, respectively, and therefore are unbiased estimates.

That is,

$$E(a) = \alpha \tag{6-18}$$

and

$$E(b) = \beta \tag{6-19}$$

since A_{xx} is a constant, $E(A_{xy}) = \beta A_{xx} + E(A_{xd})$, $E(A_{xd}) = 0$, and $E(a) = E(\bar{y} - b\bar{x}) = \alpha + \beta\bar{x} - \beta\bar{x} = \alpha$ where

$\bar{x} = \Sigma x/n$ = mean of x
$\bar{y} = \Sigma y/n$ = mean of y.

Under these assumptions, the following can also be proven:

$$\text{Var}(b) = \sigma_b^2 = E(b - \beta)^2 = \left(\frac{n}{A_{xx}}\right)\sigma_d^2 = \left(\frac{n}{A_{xx}}\right)\sigma^2 \tag{6-20}$$

and

$$E(A_{xd}^2) = n\sigma^2 A_{xx} \tag{6-21}$$

*To determine the goodness of fit of the line, many texts advocate—based on this equation—the use of $R^2 = 1 - S^2_{y_x}/S^2_y$ since, when R^2 is near unity, the variance of residuals is near zero and a "good fit is obtained" for the overall line.

$$\text{Var}(a) = \sigma_a^2 = E(a-\alpha)^2 = E(\bar{y} - b\bar{x} - \alpha)^2 = \frac{\sigma^2}{n} + \frac{n\bar{x}^2\sigma^2}{A_{xx}} = \frac{\sigma^2 \Sigma x^2}{A_{xx}} \tag{6-22}$$

the expectation of the cross-product term vanishing. Finally, the expectation of Eq. 6-14 is

$$E(S_{y_x}^2) = \sigma_d^2 = \sigma^2. \tag{6-23}$$

Since the x's are free of error under the assumptions, it can be seen from Eqs. 6-10 and 6-11 that a and b are both linear functions of the errors, d_i.

Eqs. 6-23, 6-22, and 6-20 give, respectively, the mean value of the computed variance of residuals $S_{y_x}^2$, which is based on $(n-2)$ degrees of freedom (df), and the variances of the estimates a and b. Thus if we assume that the errors d_i are normally distributed—and since $S_{y_x}^2 = S^2$ is an estimate of σ^2 based on $(n-2)$ df—then for independence of the d_i, and b and S, we have that

$$t_b = t(n-2) = \frac{(b-\beta)\sqrt{A_{xx}}}{S\sqrt{n}} \tag{6-24}$$

follows Student's t distribution with $(n-2)$ df. Hence Eq. 6-24 can be used for testing the hypothesis that $\beta = 0$ or that the true slope β equals any other constant value β_0 we may choose. Moreover, a confidence bound on the true unknown value of β may be found from Eq. 6-24.

The customary test of significance for the intercept widely used in textbooks on statistics is—in a manner similar to Eq. 6-24—given by

$$t_a = t(n-2) = \frac{(a-\alpha)\sqrt{A_{xx}}}{S\sqrt{\Sigma x_i^2}} = \frac{a-\alpha}{S\sqrt{1/n + n\bar{x}^2/A_{xx}}} \tag{6-25}$$

which follows Student's t distribution with $(n-2)$ df under the null hypothesis. Futhermore, a confidence bound is found on the true unknown intercept α from Eq. 6-25. The use of Eq. 6-25 in this connnection is quite proper if, before examining the data, we decide in advance to use the t test for a hypothesized value of α in Eq. 6-25 or to place a confidence bound on the true unknown intercept α. It is also proper if we intend to place confidence bounds on $\eta_0 = \alpha + \beta x_0$ for selected x_0, in which case we would replace a in Eq. 6-25 by $a + bx_0$, α by $\alpha + \beta x_0$, and $\bar{x}$ by $(x_0 - \bar{x})$. However, if we make multiple statements about the *line* by picking several or many values of x, then $t_{\gamma/2}(n-2)$ must be replaced by $\sqrt{2F_\gamma(2, n-2)}$, where $F_\gamma(2, n-2)$ is the upper γ probability level of Snedecor's F with 2 and $(n-2)$ df. Here the probability is now $\geq 1-\gamma$ that all such statements are simultaneously correct. The reader is referred to Scheffé (Ref. 3). Thus if a confidence bound on α is one of many such statements, one should use, instead of Eq. 6-25,

$$a \pm \sqrt{2F(2,n-2)}(S)\sqrt{1/n + n\bar{x}^2/A_{xx}} \tag{6-26}$$

where $F(2,n-2)$ follows the Fisher-Snedecor F distribution with 2 and $(n-2)$ df.

If we pick some values of x, say x^* (including $x=0$) and substitute this value of $x = x^*$ into the equation of the fitted line, i.e., into $y = a + bx^*$, then all confidence bounds desired may be found from Eq. 6-26 by replacing a by $a + bx^*$, the $\bar{x}^2$ under the radical by $(x^* - \bar{x})^2$, and proper selection of the percentage point of F by using Scheffé's theorem (Ref. 3).

To test the joint hypothesis that $\alpha = \alpha_0$ and $\beta = \beta_0$, we use the F distribution with 2 and $(n-2)$ df, i.e.,

$$F_0(2, n-2) = [n(a-\alpha_0)^2 + 2n\bar{x}(a-\alpha_0)(b-\beta_0) + (\Sigma x^2)(b-\beta_0)^2]/(2S^2). \tag{6-27}$$

A joint confidence *region* on α and β may be found from Eq. 6-27 by determining various pairs of α_0 and β_0 for which Eq. 6-27 gives the values of F not exceeding the selected confidence level $F_\gamma(2, n-2)$.

A confidence region on any number of future values of y for given values $x = x_0$ may be found from

$$a + bx_0 + \sqrt{2F(2, n-2)}(S)\sqrt{1 + 1/n + n(x_0 - \bar{x})^2/A_{xx}} \tag{6-28}$$

where we have simply added the variance of an individual, i.e., the factor one under the last radical of Eq. 6-28.

Example 6-1:

Given the data of Table 6-1, fit a line for the gravimetric determination of calcium on the values x actually present; find the standard error of residuals, and test whether the slope $\beta = 1$ and the intercept $\alpha = 0$.

Using the data of Table 6-1, we calculate the following:
$n = 10, \Sigma x = 311, \Sigma x^2 = 10{,}100, \bar{x} = 31.10, S_x = 6.90, A_{xx} = 4279, \Sigma y = 310.10, \Sigma y^2 = 10{,}055.09, \bar{y} = 31.01, S_y = 6.98, A_{yy} = 4388.89, \Sigma xy = 10{,}074.80, A_{xy} = 4306.90, \sqrt{S_{xy}} = \sqrt{A_{xy}/[n(n-1)]} = 6.92, b = A_{xy}/A_{xx} = 1.0065, a = \bar{y} - b\bar{x} = -0.2922, S^2_{y_x} = (A_{yy} - A^2_{xy}/A_{xx})/[n(n-2)] = 0.6739$, and $S_{y_x} = 0.8209$. As already indicated, we are particularly interested in whether the true slope of the line is 45 deg ($\beta = 1$) and whether the true intercept can be considered to be zero, indicating proper calibration for the gravimetric determination (new) method. To test whether $\beta = 1$, we compute t_b from Eq. 6-24

$$t_b = (1.0065 - 1.0000)\sqrt{4279}/[(0.8209)\sqrt{10}] = 0.16$$

which is not statistically significant at the 95% level. To test whether $\alpha = 0$, we compute t_a by Eq. 6-25,

$$t_a = (0.2992 - 0)/\{0.8209[1/10 + 10(31.1)^2/4279]^{1/2}\} = -0.23$$

which is not significant either. Hence we conclude the slope is unity and the calibration also is correct for $n = 10$ items.

To make the joint test of hypothesis that $\alpha = 0$, $\beta = 1$, we use Eq. 6-27 and find that the observed $F(2, n-2) = F(2,8) = 0.074$, which is not significant at the 95% level; we, therefore, conclude that the line is indeed a good fit to the data.

For any given level of Ca0 actually present, such as $x = x^* = 20$, or 40, the standard error of prediction for that value from the fitted line, $y = a + bx = -0.2922 + 1.0065x^*$, is given by

$$S_{y_x}\sqrt{1/n + n(x^* - \bar{x})^2/A_{xx}}. \tag{6-29}$$

Thus if we take $x^* = 20$ and substitute this value in Eq. 6-29 of the fitted line, we get its standard error

$$S_y \text{ (predicted)} = 0.8209\sqrt{1/10 + 10(20 - 31.1)^2/4279} = 0.51 \text{ mg.}$$

As already indicated, the confidence interval for a future (individual) observation y_0 on y, corresponding to a given true value of $x = x_0$, may be found from Eq. 6-28*. Thus a 95% confidence bound on a new observed y for $x = x_0 = 20$, $t_{0.975}(8) = 2.306$, is given by

$$-0.2922 + 1.0065(20) \pm t_{0.975}(8)\,(0.8209)\sqrt{11/10 + 10(20 - 31.1)^2/4279}$$

$$= 19.84 \pm 2.23 = 17.61 \text{ to } 22.08 \text{ mg.}$$

(Note that the standard error for the single future observation is 0.97 compared to the value of only 0.51 mg based on the same point substituted into the equation of the fitted line.)

Since x is regarded as the "true" value, measured or determined without error, then of more particular interest might be confidence bounds on the true amount of Ca0 for a given measurement by the (new) gravimetric method. Thus suppose we have measured y to be $y = y' = 20.1$ mg, then the approximate

*With $\sqrt{2F}$ replaced by t for a particular *a priori* value of $x = x_0$.

confidence bound on x, obtained by substituting y' in the equation of the line $y' = a + bx$ and solving for x, may be found for the *a priori* y' from

$$(y' - a)/b + t_{\gamma/2}(n-2)\,(S/b)\,\sqrt{1/n + n[(y'-a)/b - \bar{x}]^2/A_{xx}}. \tag{6-30}$$

Thus for $y' = 20.1$, substitution in Eq. 6-30 in which $t_{\alpha/2}(n-2) = 2.306$, gives a confidence bound on x of $20.26 \pm 1.15 = 19.11$ to 21.41, so that the appropriate probability statement on x for $y' = 20.1$ mg is

$$Pr[19.11 \text{ mg} \leq \text{True Ca0} \leq 21.41 \text{ mg}] = 0.95.$$

Note that we have used the fitted line to improve the accuracy of prediction, as compared to that of a single determination, by the new method. If the error of prediction is too large for the practical problem involved, we might improve on precision by taking more points (especially at the ends for a fitted line) or by concluding that a better measurement method is needed.

Finally, concerning the example, we did not have a physical law or hypothesis for the fitted equation. Therefore, we had to use the line. In some of the later examples in this chapter, we will consider functional relationships or appropriate physical laws in our analyses.

At this particular point it is interesting to use the two-instrument model of Chapter 2 and to estimate the standard deviation of the errors in determining both x and y. In this connection, the variance of the errors in the determination (or measurement) of y with the new method is

$$S_y^2 - S_{xy} = 4388.9/90 - 4306.90/90 = 0.911$$

or the

$$\text{est}\sigma_{ey} = \sqrt{0.911} = 0.95 \text{ mg}$$

where

$\text{est}\sigma_{ey}$ = estimate of the standard error of measurement of y (first subscript means error).

On the other hand, the variance of the errors in the determination of x, assuming the two-instrument model of Chapter 2, is

$$(4279 - 4388.89)/90 < 0$$

which is negative. Thus we must conclude that $\sigma_{ex} = 0$, or the errors of measurement in the determination of x is indeed zero, as was assumed at the start.

6-2.3 USE OF DEVIATIONS FROM THE MEAN

Suppose that instead of fitting the line $y = a + bx$, we had fitted $y = a_0 + (x_i - \bar{x})$, i.e., measure each x_i from its mean. In this case, our normal equations become

$$na_0 + [\Sigma(x_i - \bar{x})]b = \Sigma y_i$$

and

$$[\Sigma(x_i - \bar{x})]a_0 + [\Sigma(x_i - \bar{x})^2]b = \Sigma(x_i - \bar{x})y_i = \Sigma x_i y_i - \bar{x}\Sigma y_i = \frac{A_{xy}}{n}\ .$$

But since

$$\Sigma(x_i - \bar{x}) = \Sigma x_i - n\bar{x} = 0, \text{ then } na_0 = \Sigma y_i \text{ or}$$
$$a_0 = \frac{1}{n}\Sigma y_i = \bar{y}\ . \tag{6-31}$$

Moreover,

$$b = \frac{\Sigma(x_i - \bar{x})y_i}{\Sigma(x_i - \bar{x})^2} = \frac{A_{xy}}{n\Sigma(x_i - \bar{x})^2} = \frac{A_{xy}}{A_{xx}}$$

which is the same as Eq. 6-11.
Note, however, that $a = a_0 - b\bar{x} = \bar{y} - b\bar{x}$, which agrees with the intercept a fitted from the equation $y = a + bx$ as before. The importance of this result is that by a simple transformation of the independent variable, i.e., by choosing the origin of the analysis for x at its mean value, we can always eliminate the constant term if desired.

In Eq. 6-25 the variance of the intercept a without transformation of data is $\Sigma x^2/A_{xx} = \sigma_d^2$. The variance of a_0, however, is σ_d^2/n, as one might surmise since it is simply the variance of an average value.

6-2.4 TRANSFORMATION OF ORIGINAL DATA FOR LINEAR LEAST SQUARES

In many problems the original observed variables x and y may be so large (or small) that it would be inconvenient to work directly with them. Hence we may want to subtract some constant from one or both variables or to multiply or divide the original numbers by some constant factor. Thus suppose we transform the x_i and y_i as follows:

$$u_i = c(x_i - h) \quad ; \qquad v_i = d(y_i - k) \tag{6-32}$$

where c, d, h, and k are selected constants, which bring about workable values, and u_i and v_i are independent variables.

Making these transformations, we find:

$$A_{uv} = n\Sigma uv - (\Sigma u)(\Sigma v) = cdA_{xy} \text{ or } A_{xy} = A_{uv}/(cd) \tag{6-33}$$

$$A_{uu} = c^2 A_{xx}, \text{ or } A_{xx} = A_{uu}/c^2 \tag{6-34}$$

$$A_{vv} = d^2 A_{yy}, \text{ or } A_{yy} = A_{vv}/d^2 \tag{6-35}$$

$$\Sigma u_i = c\Sigma x_i - nch \quad ; \quad \Sigma v_i = d\Sigma y_i - ndk. \tag{6-36}$$

Hence the slope b becomes

$$b = \frac{A_{xy}}{A_{xx}} = \frac{A_{uv}}{cd} \cdot \frac{c^2}{A_{uu}} = \frac{c}{d} \cdot \frac{A_{uv}}{A_{uu}} \tag{6-37}$$

and the intercept a is then

$$\begin{aligned} a &= \frac{1}{n}(\Sigma y_i - b\Sigma x_i) = \frac{1}{nd}\left[\Sigma v_i + ndk - \frac{A_{uv}}{A_{uu}}(\Sigma u_i + nch)\right] \\ &= \frac{\bar{v}}{d} - \frac{A_{uv}\bar{u}}{A_{uu}d} + k - \frac{chA_{uv}}{dA_{uu}}. \end{aligned} \tag{6-38}$$

The variance of residuals S^2 will be affected only by the scale constant d, i.e.,

$$S^2 = \frac{1}{n(n-2)d^2}\left(A_{vv} - \frac{A_{uv}^2}{A_{uu}}\right). \tag{6-39}$$

The SS on the original scale becomes

$$\Sigma x_i^2 = \left(\frac{1}{c^2}\right)\Sigma u_i^2 + \left(\frac{2h}{c}\right)\Sigma u_i + nh^2. \tag{6-40}$$

Therefore, by using these equations, we may work with the transformed variables u and v and find the required parameter estimates for the original variables x and y. Indeed, such transformations are often very convenient or necessary in regression analysis calculations.

6-2.5 EQUAL SPACING OF THE INDEPENDENT VARIABLE

In some problems it may be that the x's are equally spaced, i.e., the x_i may be represented algebraically as

$$x_1 = e\,;\; x_2 = e + f\,;\; x_3 = e + 2f,\ldots;\; x_i = e + (i-1)f,\ldots;$$

and

$$x_n = e + (n-1)f \tag{6-41}$$

where f is the width of the uniform interval and e is a convenient origin. In this case, it can be shown that

$$\sum_{i=1}^{n} x_i = ne + \frac{n(n-1)}{2}f \tag{6-42}$$

$$\sum_{i=1}^{n} x_i^2 = ne^2 + 2ef\frac{n(n-1)}{2} + f^2\left(\frac{n-1}{6}\right)(n)(2n-1) \tag{6-43}$$

$$A_{xx} = \frac{n^2f^2}{12}(n^2-1) \tag{6-44}$$

and

$$A_{xy} = \frac{nf}{2}\sum_{i=1}^{n}(2i-n-1)y_i. \tag{6-45}$$

Hence for the slope b we obtain

$$b = \frac{A_{xy}}{A_{xx}} = \frac{6\sum_{i=1}^{n}(2i-n-1)y_i}{nf(n^2-1)} \tag{6-46}$$

and for the intercept a we have

$$a = \frac{1}{n}(\Sigma y_i - b\Sigma x_i) = \frac{1}{n}\left\{\Sigma y_i - \left[\frac{6e - 3f(n-1)}{f(n^2-1)}\right]\sum_{i=1}^{n}(2i-n-1)y_i\right\} \tag{6-47}$$

and finally the variance of residuals S^2 is

$$S^2 = \frac{1}{n(n-2)}\left(A_{yy} - \frac{A_{xy}^2}{A_{xx}}\right)$$

$$= \frac{1}{n(n-2)}\left\{A_{yy} - \frac{3\left[\sum_{i=1}^{n}(2i-n-1)y_i\right]^2}{n^2-1}\right\}. \tag{6-48}$$

Hence for equal spacing of the independent variable x, the key equations involve the y's, e, f, and n. These equations give all the information required to find the values of $\hat{\sigma}_a^2$, $\hat{\sigma}_b^2$, t_a, t_b, etc., as needed.

Although we have dealt with x and y to the first power, either or both variables may be more complicated as we will see in the sequel.

Although par. 6-3 relates to a special case of linear regression, it is frequently applied in the physical sciences and indeed in many Army problems.

6-3 LINEAR REGRESSION AND FUNCTIONAL RELATIONS—BOTH VARIABLES SUBJECT TO ERROR, BUT INDEPENDENT VARIABLE CONTROLLED

6-3.1 PRELIMINARIES TO ESTABLISH "FREE OF ERROR" IN INDEPENDENT VARIABLE

The problem of fitting lines or linear functional relations of some physical significance becomes much more complex for the important case in which both the dependent and independent variables are subject to (random) measurement error. Here, one has the problem of finding the physical or functional relation for the true unknown parts of x and y in the presence of "noise", and it clearly becomes important to have some knowledge of, or to be able to estimate, the relative sizes of the errors in y as compared to those in x, whether these errors are correlated with each other, or whether errors of measurement in the variables depend on the magnitude of physical values studied, etc. Indeed, there are more parametric quantities of interest than can possibly be estimated without rather severe assumptions on what may actually be happening. The reader will appreciate this in what follows; however, it will be instructive to first return to the data of Table 6-1 and Eq. 6-1 to check our assumptions in the analysis of that data. In particular, we assume that x, the amount of CaO actually present, was "free of error" and further "verified" this with the aid of the principles of Chapter 2. However, let us now pursue an allied, but somewhat different, analysis. In this connection, suppose we now replace Eqs. 6-1 and 6-2 by the model

$$x_i = \mu_i + e_i \tag{6-49}$$

and

$$y_i = \alpha + \beta\mu_i + d_i = \eta_i + d_i. \tag{6-50}$$

In other words, x is not now (as) free of error but is measured with (random) error e; in addition, y has error d as before, so that our problem is to estimate the true, but unknown, relation $\eta = \alpha + \beta\mu$, which is "covered" with noise. μ is not considered a random variable here, but rather a mathematical variable or a physical one (a "controlled" variable, i.e., purposely varied).

In the analysis of par. 6-2, we considered that the errors e_i were zero, or quite inconsequential, and that the variance of errors was zero, i.e., $\sigma_e^2 = 0$. For the observed x_i in Eq. 6-49, we have from the definitions of variances and covariances that

$$S_x^2 = \Sigma(x_i - \bar{x})^2/(n-1) = S_\mu^2 + 2S_{\mu e} + S_e^2. \tag{6-51}$$

Likewise, for the observed y_i in Eq. 6-50, we have

$$S_y^2 = \beta^2 S_\mu^2 + 2\beta S_{\mu d} + S_d^2 = S_\eta^2 + 2S_{\eta d} + S_d^2 \tag{6-52}$$

and for the covariance between the observed x's and y's, we have

$$S_{xy} = \beta S_\mu^2 + S_{\mu d} + \beta S_{\mu e} + S_{de}. \tag{6-53}$$

For the hypothesized or true linear relationship, $\eta = \alpha + \beta\mu$, we must be able to estimate α and β accurately from the data. The expected values of S_d^2 and S_e^2 are σ_d^2 and σ_e^2, respectively, i.e., the variances in errors (of measurement) of y and x, and the quantity $S_\mu^2 (= \sigma_\mu^2$ also), or S_μ, is a measure of the variation over the range of interest of the experiment. It is certainly important to know something about the relative magnitudes of σ_d, σ_e, and σ_μ for such information is, in fact, needed for the best estimates of α and β. Finally, the problem is made

more difficult because of the covariances $S_{\mu d}$, $S_{\mu e}$, and S_{de}, which could have nonzero expectations equal to $\sigma_{\mu d}$, $\sigma_{\mu e}$, and σ_{de}, respectively, in some applications. Thus we have the formidable problem of being interested in eight parameters—α, β, σ_d^2, σ_e^2, σ_μ^2, $\sigma_{\mu d}$, $\sigma_{\mu e}$, and σ_{de}—and having far too few conditions from which to estimate them! By assuming that the errors are not correlated with each other or with the levels of the values taken by μ and that they have constant variance over the range, the expectations of all the covariance terms vanish, and we are left with the expectations of Eqs. 6-51, 6-52, and 6-53, which are

$$\sigma_x^2 = \sigma_\mu^2 + \sigma_e^2 \tag{6-54}$$

$$\sigma_y^2 = \beta^2\sigma_\mu^2 + \sigma_d^2 \tag{6-55}$$

and

$$\sigma_{xy} = \beta\sigma_\mu^2. \tag{6-56}$$

Even though α is absent from these three equations, we still have four unknowns—β, σ_d^2, σ_e^2, and σ_μ^2. Thus it is quite evident that some knowledge, even from past experience of the relative sizes of the variances in errors, σ_d^2 and σ_e^2, becomes critical indeed. If we know for the problem at hand $\sigma_d = \sigma_e$, solutions are forthcoming (although from small samples we could still run into negative estimates of the variances). With this background, however, we may proceed with the analysis of the data of Table 6-1 and later discuss needed aspects of the overall problem of estimation.

For the example of Table 6-1, we found that $b = 1.0065$ for the estimate of β and that this value did not depart significantly from unity. Thus since $S_{xy} = A_{xy}/[n(n-1)]$, we might estimate σ_μ^2 from equation Eq. 6-56, i.e., from $S_{xy}/b = 47.85/1.0065 = 47.54$, (or even from $S_{xy}/1 = 47.85$), and σ_e^2 from Eq. 6-54. We get $\hat{\sigma}_e^2 = S_x^2 - \hat{\sigma}_\mu^2 = 47.54 - 47.54 = 0$, so our assumption that $\sigma_e = 0$, or that x is "free of error" (except for possible calibration bias), certainly seems valid for the analysis of Table 6-1 data. We are therefore confident in treating x as "free of error", as we did. Hopefully, this makes clear what we mean by "free of error".

6-3.2 THE CONCEPT OF A CONTROLLED INDEPENDENT VARIABLE

Next, in approaching the possibility of error in both variables, we proceed with a very important result from Berkson (Ref. 4), which has a profound effect on regression problems in the physical sciences. Berkson's result states that if the independent variable x is "controlled", even though it is otherwise "measured with error", the ordinary least squares estimate of the slope in Eq. 6-11, i.e., $b = A_{xy}/A_{xx}$, gives an unbiased estimate of β for the linear fit, and $a = \bar{y} - b\bar{x}$ is also an unbiased estimate of α. To appreciate this result, we first note that so far we have considered only the errors d_i and e_i to be random variables, which have zero means, and variances σ_d^2 and σ_e^2, respectively. We have not yet considered the possibility that μ_i could be of a random character because in the physical sciences there are so many cases of interest in which random sampling with respect to the μ_i is not carried out—i.e., the x_i are varied systematically over some particular range of interest in the experiment. This being the case, the x_i are brought to nearly fixed, or "controlled", levels by setting the dial of an instrument, presetting the time or distance measurement, etc., or aiming for a fixed, or preset, level, which is measured as x_i. Thus from Eq. 6-49 we have as before that e_i is a random variable but also the μ_i has been in effect made to be random about x_i by controlling the x_i. Hence $\mu_i = x_i - e_i$, and, upon substituting this relation in Eq. 6-50, we have

$$y_i = \alpha + \beta x_i + (d_i - \beta e_i). \tag{6-57}$$

But since the expectations of d_i and e_i are zero and x_i is fixed or controlled, we have the problem of fitting $y_i = \alpha + \beta x_i +$ (a random error), which reduces to that of par. 6-2, so that the ordinary least squares slope b becomes an unbiased estimate of the true and unknown slope β! This means that because of the imposed method of sampling or taking the data, we have controlled the μ_i to narrow random ranges about the selected or set x_i, which are brought to given levels, so that linear regression with error only in the dependent variable is still appropriate. Moreover, since the expectations of the errors are zero and that of b is equal to β, $a = \bar{y} - b\bar{x}$ is an unbiased estimate of the intercept α as well! Berkson's (Ref. 4) result is, therefore, of great importance in

wide fields of scientific investigation and experimentation since (1) relatively the variance in errors of x, or σ_e^2, is small compared to the overall variance of the μ_i (made possible by varying and controlling the x_i over a suitable range) and (2) the measured x_i consequently average out over the imposed range to give an unbiased estimate of β anyway. In summary, therefore, we are fortunate indeed for a wide class of problems in which we can simply ignore the errors in the independent variable. (The experience in Army research and development (R&D) is that controlling the independent variable is very widely practiced in curve fitting problems, and one infrequently encounters the case in which the μ_i are random or statistical variates except in the narrow range about the controlled x_i previously discussed. Hence the Berkson model has very wide application.) Finally, as will be seen, we may still estimate the values of the variances in errors of x and y, i.e., σ_e^2 and σ_d^2, respectively; however, the most critical problem is estimating β accurately.

In view of the Berkson development, we will give an example in penetration mechanics, the data for which we are indebted to Mr. Chester Grabarek of the Terminal Ballistics Division, US Army Ballistics Research Laboratories (USA BRL). Furthermore, the data are not linear, but lie on the branch of a hyperbola, so that we will transform the variables to near linearity for analysis and also will attempt to illuminate our analysis with some physical meaning or functional relationship.

The data are given in Table 6-2, covering an experiment on striking velocities and residual velocities for a 27-g penetrator fired at 0.5-in. armor plate.

TABLE 6-2

STRIKING VELOCITIES, RESIDUAL VELOCITIES, AND RESIDUAL MASSES FOR 27-g PROJECTILES FIRED AGAINST 0.5-in. ARMOR PLATE

Striking Velocity V_S, ft/s	Residual Velocity V_R, ft/s	Residual Mass M_R, g	$y = V_R^2/10^6$	$x = V_S^2/10^6$
2487	0	—	0	6.185
2508	0	—	0	6.290
2611	0	—	0	6.817
2631	0	—	0	6.922
2680	950	14.267	0.903	7.182
2732	1102	16.572	1.214	7.464
2735	1154	14.204	1.332	7.480
2718	1265	12.527	1.600	7.388
2646	1273	11.816	1.621	7.001
2707	1292	12.276	1.669	7.328
2846	1648	18.419	2.716	8.100
3023	2036	18.894	4.145	9.139
3051	2157	16.064	4.653	9.309
3331	2522	17.970	6.360	11.096
3579	2859	19.604	8.174	12.809
3971	3382	19.627	11.438	15.769
4274	3702	19.837	13.705	18.267

Striking velocities and residual velocities are plotted on Fig. 6-1. For the higher striking and residual velocities at the upper part of the curve, the slope should approach unity (angle of 45 deg), whereas it becomes infinite at the value of V_S for which $V_R = 0$. For the higher striking velocities, all rounds penetrate the plate until the knee of the curve is reached, at which point the chance of complete penetration varies from nearly 100% down to zero or near zero percent at the "limit" or "critical" striking velocity for which the residual or exit velocity is zero, i.e., partial penetration. In this particular problem, one is very interested in fitting an appropriate curve or law so that not only can he estimate but also place confidence bounds on the limit or critical striking velocity (x intercept). Although one might be tempted to exclude the V_S for the four cases

where $V_R = 0$, i.e., the partial penetrations, these are nevertheless valid points and will be included in our least squares analysis procedure.

A plot of the square of the residual velocities versus the square of the striking velocities (last two columns of Table 6-2) indicates a nearly linear relationship. Therefore, we will analyze the transformed variables $y = V_R^2/10^6$ and $x = V_S^2/10^6$. Also since the independent variable may for practical purposes be regarded as a controlled variable, we may treat it as being essentially "free of error" by using Berkson's model, and moreover it seems natural to regard any function of the residual velocity V_R as the dependent variable.

For the transformed variables x and y we obtain

$n = 17$, $\Sigma x = 154.546$, $\Sigma x^2 = 1598.068$, $A_{xx} = 3282.690$
$n = 17$, $\Sigma y = 59.530$, $\Sigma y^2 = 484.163$, $A_{yy} = 4686.950$
$\Sigma xy = 770.092$, $A_{xy} = 3891.441$

$b = A_{xy}/A_{xx} = 1.185$, $a = \bar{y} - b\bar{x} = 3.502 - (1.185)9.091 = -7.271$. Therefore, substitution into $y = a + bx$ yields

$$V_R^2 = 1.185 V_S^2 - 7{,}271{,}000.$$

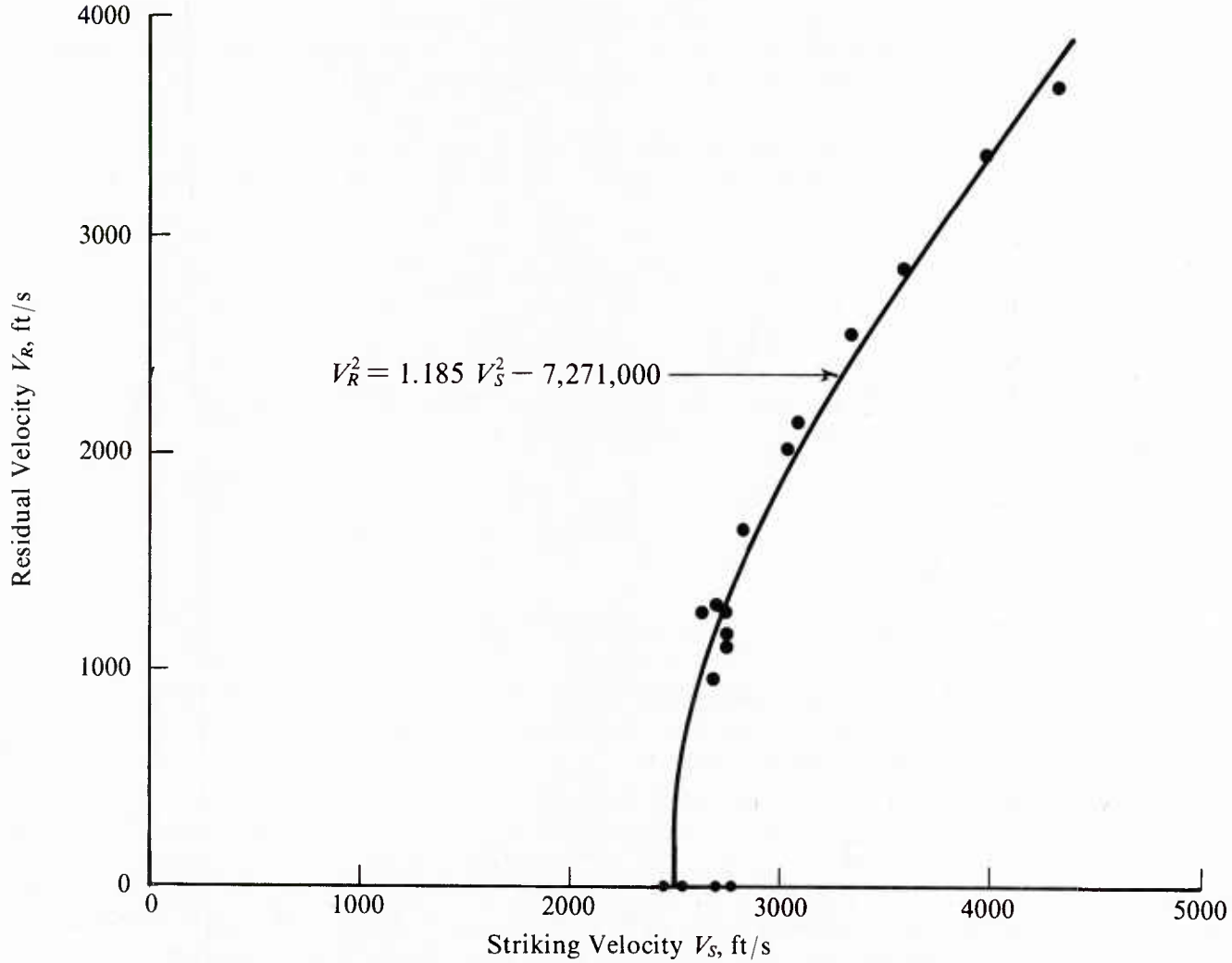


Figure 6-1. Residual Velocity vs Striking Velocity of Projectiles

When $V_R = 0$, $V_S = 2477$ ft/s, the estimated "limit" velocity. The variance of residuals is

$$S^2_{y_x} = (A_{xx}A_{yy} - A^2_{xy})/[n(n-2)A_{xx}] = 0.290, \text{ or } S_{y_x} = 0.538.$$

95% confidence bounds on the true unknown "limiting" x, i.e., for $y = 0$, are obtained from Eq. 6-30, where $y' = 0$, or that is, from

$$-a/b \pm t_{\gamma/2}\,(n-2)\,(S_{y_x}/b)[1/n + n(-a/b - \bar{x})^2/A_{xx}]^{1/2}. \tag{6-58}$$

This gives for $t_{\alpha/2}(15) = 2.131$

$$Pr[5.824 \leq x_{\text{limit}} \leq 6.448] = 0.95$$

and since $V_S^2/10^6 = x$, we have for the original data that

$$Pr[2413 \text{ ft/s} \leq V_{\text{limit}} \leq 2539 \text{ ft/s}] = 0.95$$

so that the 95% confidence bound on the true unknown limit or critical velocity is $2539 - 2413 = 126$ ft/s wide for the V_S intercept.

Had the previous statement been one of many similar ones about confidence bounds for various points on the line, Student's $t_{\gamma/2}(n-2)$ should be replaced by $\sqrt{2F_\gamma(2,n-2)}$, using the upper level of the Snedecor F, and the resulting confidence bounds for V_{limit} would be $2396 - 2555$ ft/s, or 159 ft/s wide, or an increase of 33 ft/s.

The variance of residuals on the transformed scale is $S^2_{y_x} = 0.290$, but since $V_R = 1000\sqrt{y}$, we have $dV_R = 500y^{-1/2}dy$, and upon squaring and taking mean values we have the variance of residuals on the original scale of V_R, which is

$$\sigma^2_{V_R} \approx (250{,}000/\bar{y})\sigma^2_{y_x} = (250{,}000/3.502)\,(0.290) = 20{,}702$$

or

$$\sigma_{V_R} = 144 \text{ ft/s} \quad \text{(for an individual value)}.$$

At this stage we might ask whether our assumption that x is "free of error" is met, or nearly so. In this connection we note from Eq. 6-56 that $\sigma^2_\mu = \sigma_{xy}/\beta$ and, hence, that

$$\hat{\sigma}^2_\mu = \text{est}\sigma^2_\mu = A_{xy}/[n(n-1)b] = 12.07.$$

Now from Eq. 6-54 we take

$$\hat{\sigma}^2 = \hat{\sigma}^2_x - \hat{\sigma}^2_\mu = A_{xx}/[n(n-1)] - \hat{\sigma}^2_\mu = 3282.69/(17)\,(16) - 12.07 = 12.07 - 12.07 = 0$$

which gives us considerable confidence in our procedure. We also observed from Eq. 6-55 that our observed estimate of σ^2_d becomes $\hat{\sigma}^2_d = 0.28$ or $\hat{\sigma}_d = 0.53$, which converted to the original scale of V_R is 141 ft/s versus the 144 ft/s previously calculated, or a good check.

In fitting the equation

$$V^2_R = 1.185V^2_S - 7{,}271{,}000$$

we merely observed that the original data fall on the branch of a hyperbola type of curve, and hence we could linearize the data (or approximately so) by working with the squares of the striking and residual velocities. But what about the possibility of a "physical" fit or law? Here we might consider fitting the residual energy versus

the striking energy. In Table 6-2, note that a third or more of the weight of the projectiles wears away in the penetration process. Nevertheless, it might make considerable sense to treat the "measured" residual energy as the dependent variable and the striking energy as the independent variable. We will actually take our new $x = m_S V_S^2/10^8 = 27V_S^2/10^8$ and new $y = m_R V_R^2/10^8$; m_R varies as given in Table 6-2. A plot of these new x's and y's indicates a nearly linear relationship. Our key computations now become
$n = 17$, $A_{xx} = 239.301$, $A_{yy} = 187.103$, $A_{xy} = 210.721$ $b = 0.8806$, $a = -1.523$
or $y = -1.523 + 0.8806x$.

By using the average of the residual masses ($\bar{m}_R = 16.314$ for the 13* penetrating rounds), we now have the equation

$$V_R^2 = -9{,}335{,}540 + 1.457V_S^2.$$

By setting $V_R = 0$ in this equation, $V_S = 2531$ ft/s. Also since $S_{y_x} = 0.078$,

$$Pr\,[2497 \text{ ft/s} \leq V_S(\text{limit}) \leq 2565 \text{ ft/s}] \approx 0.95.$$

Thus by using the "physical" law, the confidence interval has a width of $2565 - 2497 = 68$ ft/s or 58 ft/s shorter than the one based on V_R^2 and V_S^2! (We note that this "law" does not fit as well as the other one at the upper end of the curve although the lower end is still of more interest. We also note that raising the "measured" residual energy and the striking energy to about the 0.90 or 0.95 power might produce a slightly better linear relationship, but this would begin to depart from physical considerations.)

For the transformed data based on striking energy and "measured" residual energy, we have from Eqs. 6-54, 6-55, and 6-56 that

$$\hat{\sigma}_\mu^2 = 0.88, \qquad \hat{\sigma}_e^2 \approx 0.00, \qquad \text{and} \quad \hat{\sigma}_d^2 \approx 0.10$$

so that the assumptions still seem sufficiently valid, and the relation between striking and residual velocities is taken as $V_R^2 = 1.457V_S^2 - 9{,}335{,}540$. Moreover, the standard deviation of the random measurement error d is easily converted to the original scale of the residual velocity V_R and is approximately

$$\sigma_{V_R} \approx 10^4 \sigma_{y_x}/(2\sqrt{m_R y}) = 60 \text{ ft/s},$$

a value much less than the value of 144 ft/s previously obtained for V_R^2 versus V_S^2.

In summary, we have demonstrated the importance of trying to seek a physical relationship, transforming the original variables to near linearity for the regression analysis, and then being able to make statistical or probability statements about the original variables of interest on the old scale.

If we knew that the slope of the line is unity from physical considerations, there would be little point in estimating it statistically, except for a check; consequently, the analysis would be much simplified. Also for more complex problems one might consider using various functions of the physical variables, which result in linearity with only the error of determination of that variable following a statistical distribution. Indeed, regression problems are not all statistical, nor are they all physical; rather they are a combination of both that may result in wider practical value and utility.

We mentioned that proper estimation of the slope β was important and that unbiased estimates are needed. As a result, Eqs. 6-54, 6-55, and 6-56 are of considerably more help than might be realized. To begin with, if $\sigma_e = 0$, we note by using Eqs. 6-54 and 6-56 that the proper estimate of

$$\hat{\beta} = A_{xy}/A_{xx}$$

*Some might argue that the four rounds that did not penetrate have "zero mass", but this would be strange to many ballisticians. These calculations are primarily for illustrative reasons.

as we established in Eq. 6-11. If σ_e is not zero but known, for example, from past data or experience, then Eq. 6-54 indicates that

$$\sigma_\mu^2 = \sigma_x^2 - \sigma_e^2$$

so that an unbiased estimate of β may be found (observing Eq. 6-56) from

$$\hat{\beta} = A_{xy}/(A_{xx} - n^2\sigma_e^2). \tag{6-59}$$

If σ_d is known, observing Eqs. 6-55 and 6-56, we see that an estimate of β is found from

$$\hat{\beta} = (A_{yy} - n^2\sigma_d^2)/A_{xy}. \tag{6-60}$$

If both σ_d and σ_e are known, from Eqs. 6-55 and 6-56 we obtain the estimate

$$\hat{\beta} = (A_{yy} - n^2\sigma_d^2)^{1/2}/(A_{xx} - n^2\sigma_e^2)^{1/2}. \tag{6-61}$$

The estimates from Eqs. 6-59, 6-60, and 6-61 are not ML estimates, but they do enjoy the property of being "consistent"—i.e., for large samples, they tend in probability toward the true unknown linear slope parameter β.

Since we have seen the importance of estimating the slope accurately and that the method of estimating it depends on the values of the (often unknown) variances in errors of measurement or determination, continuing knowledge of the precision of measurement of instruments—i.e., their capacity for repeatability, reproducibility, and also accuracy—becomes critical indeed. In fact, any worthwhile experiment could be planned and carried out more appropriately with such continuing knowledge of instrument precision capability since this would lead to improved analyses and predictions for the data taken. Moreover, we now see from the discussion and examples that the matter of trying to find even some linear relationship between true values of the variables studied can become complex.

In our account we have not exhausted the methods of estimating the slope β. In fact, we should mention that for the linear relation and error in both variates, grouping methods, such as that of Wald-Bartlett (Refs. 5 and 6), might be used to advantage. Grouping methods were developed primarily for the case in which the μ_i are random variables (discussed further later), but they may also be used for the case in which they are varied systematically by the investigator over particular ranges of interest. The Wald-Bartlett method for estimating β involves dividing the data ordered in the x-direction into three approximately equal groups; computing the mean x's and y's of the two extreme groups, i.e., $(\bar{x}_1,\bar{y}_1)$ and $(\bar{x}_3,\bar{y}_3)$; and estimating the slope β from

$$\hat{\beta} = (\bar{y}_3 - \bar{y}_1)/(\bar{x}_3 - \bar{x}_1). \tag{6-62}$$

(Of course, totals could be used in place of averages.) To illustrate the measured energy versus striking energy fit, we will use the top five and bottom five points and compute $m_R V_R^2/10^8$ and $27V_S^2/10^8$ for each point. This results in the following estimate of slope:

$$\hat{\beta} = \frac{(2.72 + 2.24 + 1.60 + 1.14 + 0.75) - (0.1288 + 0 + 0 + 0 + 0)}{(4.93 + 4.26 + 3.46 + 2.99 + 2.51) - (1.67 + 1.70 + 1.84 + 1.87 + 1.94)} = 0.91$$

whereas from the linear least squares fit we obtained $b = 0.88$, which indicates rather good agreement (although it does distribute the error to the independent variable, which indicates the extreme sensitivity involved).

We will not discuss the best methods of grouping and the various ramifications of the technique but will refer the reader instead to papers of Wald (Ref. 5), Bartlett (Ref. 6), Madansky (Ref. 7), and Neyman (Ref. 8).

For the case of error in both variables, we will mention finally an estimate of β that seems intuitive on practical grounds. This involves finding the slope by least squares from the linear regression of the "dependent" variable y and averaging this with the reciprocal of the slope obtained by finding the regression of x and y since both contain error. From the former we have that $b_{y_x} = A_{xy}/A_{xx}$, and from the latter that $b_{x_y} = A_{xy}/A_{yy}$. Using the preceding data, we obtain

$$b_{y_x} = 210.721/239.301 \qquad \text{and} \qquad b_{x_y} = 210.721/187.103$$

$$= 0.8806 \qquad\qquad = 1.1262$$

so that

$$\hat{\beta} = (0.8806 + 1/1.1262)/2 = 0.8843.$$

Moran (Ref. 9) treats this type of estimate.

6-4 LINEAR LEAST SQUARES WITH BOTH VARIABLES SUBJECT TO ERROR AND BOTH VARIABLES RANDOM

In this case the model of Eqs. 6-49 and 6-50 still applies, but instead of being a controlled or fixed variable, μ is now random.* (There are some problems in the physical sciences or ballistics technology that fall into this category, but we believe the controlled variable case takes priority.) The errors d_i and e_i are again considered to be normally distributed with zero means and variances σ_d^2 and σ_e^2 as before. It is easy to see that many of the equations developed in par. 6-3 still apply to the case of μ being randomly distributed. In fact, Eqs. 6-54, 6-55, 6-56, 6-59, 6-60, and 6-61 apply without alteration. It is very desirable for applications in the physical sciences that the variances in errors of measurement σ_d^2 and σ_e^2 be small compared to the variance in μ or σ_μ^2 to guarantee sufficient precision of measurement.

Although, as mentioned, we will not delve very deeply into this particular case—since the use of the controlled variable is widely practiced in the physical sciences—we will nevertheless establish a few principles of interest and record them here.

To begin with, if σ_d^2 and σ_e^2 are both known, Eq. 6-61 becomes the ML estimate of the slope β because then Eqs. 6-54, 6-55, and 6-56 are the basic ML estimates. We also see from these same equations that if σ_d^2 and σ_e^2 are both known, this case becomes an overidentified situation since actually we need to know only the ratio $\lambda = \sigma_d^2/\sigma_e^2$. In fact, if the ratio λ is known, Madansky (Ref. 7) shows that the proper estimate of β is given by

$$\hat{\beta} = \frac{A_{yy} - \lambda A_{xx} + [(A_{yy} - \lambda A_{xx})^2 + 4\lambda A_{xy}^2]^{1/2}}{2A_{xy}}. \tag{6-63}$$

This estimate of β also may be applied to the controlled independent variable case. For example, if we use the data for striking energy and measured residual energy previously discussed and assume $\lambda = 1$, we have

$$\hat{\beta} = \frac{187.103 - 239.301 + [(187.103 - 239.301)^2 + 4(210.721)^2]^{1/2}}{2(210.721)}$$

$$= 0.884$$

which is the same as the estimate from $(b_{y_x} + 1/b_{x_y})/2 = 0.884$.

Madansky (Ref. 7) gives a rather detailed discussion of the case in which the μ_i are random variables and includes grouping methods for estimating β et al.

*For improved clarity we could replace μ by x when it is a random variable. However, we believe the reader will easily grasp the proper concept when μ is used.

For a case where the μ_i are random and it is known that the slope $\beta = 1$, Grubbs (Ref. 10) gives methods for estimating the variances in the errors of measurement of x and y, i.e., techniques for estimating σ_e^2 and σ_d^2. In fact, this particular model becomes the subject of the two-instrument case of Chapter 2. We see this easily by examining Eqs. 6-49 and 6-50 in which, for a slope of unity, the quantity α simply amounts to a constant shift for the y values so that the model is the same as for the two-instrument precision estimation case. In summary, we see, therefore, that the models for linear regression and the problem of estimating the precision of measurement of (two) instruments are very closely allied.

Having covered these allied topics, indicating especially the importance of estimating the needed components of variance in both the linear regression models and the problem of estimating precision of measurement, we turn our attention to biases in estimation due to errors of determination of the independent variable.

6-5 BIASES IN ESTIMATION AND BIASES IN SIGNIFICANCE TESTS DUE TO ERRORS IN THE INDEPENDENT VARIABLE

When the independent variable x for the linear regression case is subject to errors of determination or measurement, the use of equations for estimation, such as Eq. 6-11 for the slope, or a significance test for the slope, such as Eq. 6-24, becomes subject to biases and hence could be somewhat misleading in correct judgments. Thus when both the dependent and independent variables are subject to errors, it may become advisable to exercise special care in estimation and significance testing procedures.

As an example of the existence of bias, consider estimation of the slope β by using Eq. 6-11 when the chosen model for the application is Eqs. 6-49 and 6-50. Here, the large sample value of the estimator b tends in probability to the ultimate value

$$\text{plim } b = \beta\sigma_x^2/(\sigma_x^2 + \sigma_e^2) \tag{6-64}$$

as, for example, may be found in Goldberger's book (Ref. 11). In other words, the sample value b will underestimate the true slope β, depending on just how large the variance in errors of the independent variable happens to be, as is noticed in Eq. 6-63. Hence unless the variance in errors σ_e^2 of the measurements of x are zero or quite small relative to the variance in the true values x, the amount of bias could be rather significant indeed. If, for example, we have that $\sigma_x = \sigma_e$, then the estimate b would approach

$$b \to \beta/2$$

which, of course, is quite a bias! Hence to keep the analysis simple, we see the desirability of keeping σ_e small or otherwise varying x over a large range of values in linear regression.

Biases occur and lead to inaccuracy in significance tests for linear regression when the independent variable x is subject to errors of determination. As an example, consider Student's t test of Eq. 6-24 for judging the null hypothesis that the slope of the fitted line is zero. Then, it can be shown, as in Bloch (Ref. 12), that the large sample value of Student's t, call it t_b, tends toward

$$t_b \to \sqrt{(n-1)}\,\beta\sigma_x^2/[(\sigma_x^2 + \sigma_e^2)\,(\sigma_d^2 + \beta^2\sigma_e^2)]^{1/2}. \tag{6-65}$$

Bloch (Ref. 12) shows that this means when there are errors in the independent variable x, Student's t tends to be too small. This results in lower probabilities of rejecting the null hypothesis that the coefficients of the imprecisely measured variables are actually zero. Hence we see that this really implies that Student's t values could often be low enough to cause one not to reject the null hypothesis when it is actually false. Thus use caution when x is subject to any significant error due to lower than true t values.

An illuminating discussion of the problem concerning the estimation of bias in the classical linear regression slope for the case in which the proper model is functional linear least squares is given by Reed and Wu (Ref. 13). For this case Reed and Wu also use the specified model of Eqs. 6-49 and 6-50, which contain errors of determination of both x and y, and cite the work of Richardson and Wu (Ref. 14), which shows that the expected value of the slope b in linear regression would depend on an exponential and hypergeometric

function although Eq. 6-63 is a sufficiently good approximation. Of perhaps further interest is that Reed and Wu (Ref. 13) give an approximate, one-sided confidence interval on the true unknown amount of bias in their Eq. 3.6, p. 411, and also discuss a "jackknifing" procedure.

Hopefully, this discussion will give the reader some useful insight into the fact that the ordinary classical linear regression procedures may lead to errors of analysis if they are applied to linear regression problems for the case in which both the independent and the dependent variables are subject to error.

6-6 A CONSISTENT ESTIMATOR OF THE SLOPE IN A LINEAR REGRESSION MODEL WITH ERRORS IN BOTH INDEPENDENT AND DEPENDENT VARIABLES

As pointed out by Eqs. 6-54 through 6-56, there are four key unknowns and only three equations available for the estimation procedure, and this is the source of much difficulty in linear regression for errors in both variables. Thus there exists a rather formidable difficulty to overcome. We also see that an additional parameter should not be introduced to complicate the problem unless the estimation of that parameter leads to a technique that not only gives an estimate of the new parameter but also includes estimation possibilities for one of the old parameters in Eqs. 6-54 through 6-56. This problem has, over the years, been given much thought, and some results of interest to the Army analyst have been achieved. For instance, Karni and Weissman (Ref. 15) have advanced the idea of using the serial correlation coefficient of lag 1 of the first order (forward) differences of the independent and dependent variables, and this procedure does lead to consistent estimators of the slope along with estimators of the variances of errors of the x and y and also the serial correlation coefficient. Thus, in effect, it provides all five estimates. However, the estimators of Karni and Weissman (Ref. 15) apply primarily to the case in which the true values μ_i are nonstochastic. When, for example, the pair (x_i, y_i) follows a bivariate normal distribution and the intercept term α of Eq. 6-50 is not zero, an underidentified situation arises again, and hence all parameters of interest cannot be legitimately estimated. Some authors have tended to circumvent this problem by relaxing the assumption of normality. The approach of Karni and Weissman in Ref. 15, on the other hand, suggests relaxing the independence assumption, namely, that the first order serial correlation ρ_1 should not be zero. Thus, for example, it might be expected that the Karni and Weissman model would apply to the two-instrument precision case discussed in Chapter 2, and indeed we will illustrate it in Example 6-1.

In order to outline the Karni-Weissman model, we are dealing with an independent variable x subject to error and a dependent variable y subject to error as usual. We will need the variances and the covariance of both x and y in our calculations. Also we will need the (forward) first order differences of each of the x and y observations. Hence we will define the symbols

$$dx_i = x_i - x_{i-1} \tag{6-66}$$

and

$$dy_i = y_i - y_{i-1} \tag{6-67}$$

for the forward first order differences and then use the usual symbols S_{dx}^2, S_{dy}^2, and S_{dxdy} to represent, respectively, the variance of the dx's, the variance of the dy's, and the covariance of the dx's and the dy's. With these definitions the key estimators for the Karni-Weissman model are

$$\hat{\sigma}_e^2 = \frac{S_x^2 S_{dxdy} - S_{dx}^2 S_{xy}}{S_{dxdy} - 2S_{xy}} \tag{6-68}$$

$$\hat{\beta} = \frac{S_{xy} - S_{dxdy}/2}{S_x^2 - S_{dx}^2/2} \tag{6-69}$$

$$\hat{\sigma}_\mu^2 = S_x^2 - \hat{\sigma}_e^2 \tag{6-70}$$

$$\hat{\rho}_1 = 1 - S_{dxdy}/(2\hat{\beta}\hat{\sigma}_\mu^2) \tag{6-71}$$

and

$$\hat{\sigma}_d^2 = S_y^2 - \hat{\beta}^2 \hat{\sigma}_\mu^2. \tag{6-72}$$

Therefore, if the assumptions of the Karni-Weissman linear regression model are justifiable, the four key parameters in which we are interested and the first order serial correlation coefficient can be estimated as shown in Example 6-1.

Example 6-1:

Return to the two-instrument precision of measurement Example 2-1 of Chapter 2 and the data of Table 2-2 for the first two instruments I_1 and I_2, i.e., r and s observations. Then treat s as the independent variable and r as the dependent variable (both measured with error) for the purpose of estimating the key linear regression parameters and as a check on Example 2-1.

Note under the assumptions of Example 2-1 the slope is expected to be unity, and also since there is no intercept to estimate, we expect the Karni-Weissman assumptions to apply with the additional assumption that perhaps the difficulty with the measurements of I_1 may relate to some serial correlation. Recall that for Example 2-1 we obtained a slightly negative variance in the errors of measurement for the instrument I_2. Of course, for the Karni-Weissman linear regression model we will, using their theory, have to estimate the slope β and then use it for estimation of some of the other parameters to see in advance that, if it is not equal to unity, there would be a different distribution of the precision of measurement parameters.

We exhibit the relevant data for this example in Table 6-3 and obtain the following pertinent calculations using only 29 observations for I_1 by deleting the value 10.01 for the corresponding lost round of I_2:

$$S_y^2 = 0.04675448 \qquad S_x^2 = 0.045112315 \qquad S_{xy} = 0.045581897 \qquad S_{dx}^2 = 0.069108995$$

$$S_{dxdy} = 0.06882328.$$

(Note in our problem there is no need to use the dy_i alone.) By using Eqs. 6-68 through 6-72, we obtain these estimates:

$\hat{\sigma}_e^2 = 0.0020296$ (which makes the second instrument less precise)

$\hat{\beta} = 1.058008$, $\hat{\sigma}_\mu^2 = 0.0430827$ (less product variability)

$\hat{\rho}_1^2 = 0.24506$ and $\hat{\sigma}_d^2 = -0.0014715$ (to be taken as zero).

We observe that with the Karni-Weissman analysis, the slope is slightly larger than unity, and this results in switching the negative variance of errors of measurement to the first instrument. Also the product variance is decreased slightly, and the second instrument is made less precise since the variance in errors for instrument I_1 seems near zero! In summary, we should say that we did not gain a great deal more understanding about our two-instrument precision of measurement problem by using the Karni-Weissman linear regression model although there could be some serial correlation in the readings of I_1, and there could be other applications to which the Karni-Weissman model would apply better.* Finally, perhaps we are trying to get too much out of the slightly different approaches! Moreover, we expect to encounter the problem of negative estimates of variances in such studies anyway.

Hopefully, this background on the linear regression problem with error in only the dependent variable on one hand, and errors in both variables on the other, may give the Army analyst sufficient background to make rather extensive applications or may lead him to further literature as needed. We now proceed to other models of interest. For example, we will discuss the fitting of planes, parabolas, and the use of orthogonal polynomials for equally spaced independent variables before finally touching upon the problem of nonlinear regression.

* We do not particularly recommend the use of grouping methods, such as the use of Eq. 6-62, because Neyman and Scott (Ref. 16) have shown that schemes based on the orders of the observations do not lead to consistent estimation.

TABLE 6-3
FUZE BURNING TIMES AND FORWARD FIRST ORDER DIFFERENCES FOR TWO INSTRUMENTS

Observer I_1		Observer I_2	
y (=r), s	dy, s	x(=s), s	dx, s
10.10		10.07	
9.98	-0.12	9.90	-0.17
9.89	-0.09	9.85	-0.05
9.79	-0.10	9.71	-0.14
9.67	-0.12	9.65	-0.06
9.89	0.22	9.83	0.18
9.82	-0.07	9.75	-0.08
9.59	-0.23	9.56	-0.19
9.76	0.17	9.68	0.12
9.93	0.17	9.89	0.21
9.62	-0.31	9.61	-0.28
10.24	0.62	10.23	0.62
9.84	-0.40	9.83	-0.40
9.62	-0.22	9.58	-0.25
9.60	-0.02	9.60	0.02
9.74	0.14	9.73	0.13
10.32	0.58	10.32	0.59
9.86	-0.46	9.86	-0.46
9.65	-0.21	9.64	-0.22
9.50	-0.15	9.49	-0.15
9.56	0.06	9.56	0.07
9.54	-0.02	9.53	-0.03
9.89	0.35	9.89	0.36
9.53	-0.36	9.52	-0.37
9.52	-0.01	9.52	0.00
9.44	-0.08	9.43	-0.09
9.67	0.23	9.67	0.24
9.77	0.10	9.76	0.09
9.86	0.09	9.84	0.08

Note: $dy = y_i - y_{i-1}$ and $dx = x_i - x_{i-1}$.

6-7 THE PLANE: ONE VARIABLE z (THE DEPENDENT VARIABLE) SUBJECT TO ERROR

In this case, we seek the relation between a dependent variable (subject to some error of determination) and two independent variables x and y, which are relatively free of error, or we seek the regression of z on x and y by the method of least squares. Also, from the physical standpoint, we are very interested in whether the fitted plane is unbiased, i.e., can be regarded as representing the functional or structural relation between the true values of z, and x and y. We will assume that the measured values of x and y are both "free of error", whereas the observed values of z are subject to a (random) error of measurement. Thus the functional relation may be represented by

$$z = \alpha + \beta x + \gamma y \tag{6-73}$$

where

α = true unknown coefficient.

The model, or assumption, considered for the observed values (x_i,y_i,z_i) is

$$x_i = \text{a variable, free of error}$$
$$y_i = \text{a variable, free of error}$$
$$z_i = \alpha + \beta x_i + \gamma y_i, \text{ subject to error } e_i \approx N(0,\sigma_e^2).$$

We propose to fit the equation

$$z = a + bx + cy \tag{6-74}$$

to the observed data by determining a, b, and c (which will be estimates of α, β, and γ, respectively) by the method of least squares, i.e., such that the SS of the deviations (observed minus fitted values) are a minimum. We have

$$\phi = \sum_{i=1}^{n}(z_i - a - bx_i - cy_i)^2 \tag{6-75}$$

to be minimum. Note that for observed means $\bar{z} = a + b\bar{x} + c\bar{y}$. Hence since the A_{uv} are not origin dependent and to simplify the algebra, we make this substitution in ϕ and obtain

$$\phi = \sum_{i=1}^{n}[(z_i - \bar{z}) - b(x_i - \bar{x}) - c(y_i - \bar{y})]^2$$

which is to be a minimum. (Note that only b and c need estimation initially.)

Differentiating with respect to b and c, we get

$$\frac{\partial\phi}{\partial b} = -2\Sigma(x_i - \bar{x})[(z_i - \bar{z}) - b(x_i - \bar{x}) - c(y_i - \bar{y})] = 0 \tag{6-76}$$

$$\frac{\partial\phi}{\partial c} = -2\Sigma(y_i - \bar{y})[(z_i - \bar{z}) - b(x_i - \bar{x}) - c(y_i - \bar{y})] = 0. \tag{6-77}$$

Solving for b, c and a, we get

$$b = \frac{A_{xz}A_{yy} - A_{yz}A_{xy}}{A_{xx}A_{yy} - A_{xy}^2} \tag{6-78}$$

$$c = \frac{A_{xx}A_{yz} - A_{xy}A_{xz}}{A_{xx}A_{yy} - A_{xy}^2} \tag{6-79}$$

$$a = \bar{z} - b\bar{x} - c\bar{y} = \frac{1}{n}[\Sigma z_i - b\Sigma x_i - c\Sigma y_i]\,. \tag{6-80}$$

The variance of residuals is given by

$$S^2 = \left(\frac{1}{n-3}\right)\sum_{i=1}^{n}[(z_i - \bar{z}) - b(x_i - \bar{x}) - c(y_i - \bar{y})]^2 \tag{6-81}$$

or

$$\text{est}\sigma_e^2 = S^2 = \left[\frac{1}{n(n-3)}\right](A_{zz} - bA_{xz} - cA_{yz})\,. \tag{6-82}$$

Under the assumption Eq. 6-73, it can be shown that the mean or expected values of a, b, and c are, respectively, α, β, and γ. Hence for the model assumed the method of least squares gives an unbiased estimate (with minimum variance) of the functional or structural relation between the true values of z and the (fixed, i.e., "free of error") variates x and y if Eq. 6-73 is the proper law.

Also by methods indicated previously for the line, it can be shown that

$$\text{est}\sigma_a^2 = \frac{nS^2[\Sigma x_i^2 \Sigma y_i^2 - (\Sigma x_i y_i)^2]}{A_{xx}A_{yy} - A_{xy}^2} \tag{6-83}$$

$$\text{est}\sigma_b^2 = \frac{nA_{yy}S^2}{A_{xx}A_{yy} - A_{xy}^2} \tag{6-84}$$

$$\text{est}\sigma_c^2 = \frac{nA_{zz}S^2}{A_{xx}A_{yy} - A_{xy}^2} \tag{6-85}$$

We now have all the information required for the usual Student's t tests to judge the hypotheses concerning whether the true parameters α, β, and γ can be regarded as being equal to zero or any selected constant values of some particular physical interest.

For example, to test whether the true slope β — in the functional or structural relation $z = \alpha + \beta x + \gamma y$—is equal to zero, we use Student's t test based on

$$t = \frac{b - 0}{\hat{\sigma}_b} = \frac{b\sqrt{A_{xx}A_{yy} - A_{xy}^2}}{S\sqrt{nA_{yy}}} \tag{6-86}$$

with $(n - 3)$ df.

Example 6-2:

The data in Table 6-4 give the ballistic limits* (BL) for various thicknesses and Brinell hardness numbers (BHN) of armor plate when tested with cal .50 armor-piercing (AP) bullets. (The plates of armor were placed at an angle of obliquity of 42 deg from the line of fire.) It is desired to find the linear regression equation of the BL z on the thickness x and BHN y.

We have

$N = 20$		
$\Sigma x_i = 4.996$	$\Sigma x_i^2 = 1.249116$	$\Sigma x_i y_i = 1837.670$
$\Sigma y_i = 7356$	$\Sigma y_i^2 = 2{,}749{,}670$	$\Sigma x_i z_i = 5900.253$
$\Sigma z_i = 23{,}583$	$\Sigma z_i^2 = 28{,}468{,}483$	$\Sigma y_i z_i = 8{,}795{,}787$
$A_{xx} = 0.022304$	$A_{xy} = 2.824$	$\bar{x} = 0.2498$
$A_{yy} = 882{,}664$	$A_{xz} = 184.392$	$\bar{y} = 367.8$
$A_{zz} = 13{,}211{,}771$	$A_{yz} = 2{,}439{,}192$	$\bar{z} = 1179.15.$

To determine the coefficients a, b, and c in $z = a + bx + cy$, we have from Eqs. 6-78, 6-79, and 6-80 that

$$b = \frac{155{,}867{,}902.08}{19{,}678.96288} = 7920.534$$

$$c = \frac{53{,}883.0154}{19{,}678.963} = 2.738102$$

*The BL of armor plate represents that striking velocity for which 50% of the projectiles penetrate the plate. BL is known to be highly variable as compared to thickness and BHN measurements.

TABLE 6-4

BALLISTIC LIMIT vs ARMOR THICKNESS AND BRINELL HARDNESS

BL z, ft/s	Thickness x, in.	BHN y
927	0.253	317
978	0.258	321
1028	0.259	341
906	0.247	350
1159	0.256	352
1055	0.246	363
1335	0.257	365
1392	0.262	375
1362	0.255	373
1374	0.258	391
1393	0.253	407
1401	0.252	426
1436	0.246	432
1327	0.250	469
950	0.242	275
998	0.243	302
1144	0.239	331
1080	0.242	355
1276	0.244	385
1062	0.234	426

and

$$a = \bar{z} - b\bar{x} - c\bar{y} = -1806.473.$$

The tentative regression equation we fit is taken as

$$BL = -1806.473 + 7920.534\ (\text{thickness}) + 2.738\ (\text{BHN}).$$

The variance of residuals is calculated to be

$$S^2 = \frac{1}{n(n-3)}\,(A_{zz} - bA_{xz} - cA_{yz}) = 14{,}919.2$$

and

$$nS^2 = 298{,}384.2.$$

Then

$$\hat{\sigma}_c^2 = \frac{nS^2 A_{xx}}{A_{xx}A_{yy} - A_{xy}^2} = 0.33819 \text{ and } \hat{\sigma}_c = 0.58154$$

$$\hat{\sigma}_b^2 = \frac{nS^2 A_{yy}}{A_{xx}A_{yy} - A_{xy}^2} = 13{,}383{,}479.26 \text{ and } \hat{\sigma}_b = 3658.344$$

$$\hat{\sigma}_a^2 = \frac{nS^2[\Sigma x^2 \Sigma y^2 - (\Sigma xy)^2]}{A_{xx}A_{yy} - A_{xy}^2} = 873{,}756.3 \text{ and } \hat{\sigma}_a = 934.749.$$

Moreover, Student's t tests of the intercept and coefficients are

$$t_a = \frac{a}{\hat{\sigma}_a} = -1.933$$

$$t_b = \frac{b}{\hat{\sigma}_b} = 2.165$$

$$t_c = \frac{c}{\hat{\sigma}_c} = 4.708.$$

Since $t_{0.05} = 2.11$ for $\nu = 17$ df, the slope b is significantly different from zero at the 5% level. The coefficient of BHN is highly significant ($P < 0.005$). Thus we would adopt the previously given equation for predicting BL from thickness and BHN under conditions similar to those of the executed test. (In this particular case, the thicknesses appear to vary randomly in character, as do the BHN to some extent. If the thicknesses had varied over a wide range, the slope b would have been highly significant.)

The variance of a value of z predicted from Eq. 6-74 is given by the following equation for any selected values x and y:

$$\sigma_z^2 = \frac{\sigma_e^2}{n} + (x - \bar{x})^2\sigma_b^2 + (y - \bar{y})^2\sigma_c^2 + 2(x - \bar{x})(y - \bar{y})\sigma_{bc}. \tag{6-87}$$

Estimates of σ_a^2, σ_b^2, and σ_c^2 are given by Eqs. 6-83, 6-84, and 6-85, whereas an estimate of σ_{bc} is given by

$$\sigma_{bc} = \frac{-nA_{xy}S^2}{A_{xx}A_{yy} - A_{xy}^2}. \tag{6-88}$$

6-8 THE PARABOLA: ONE VARIABLE z (THE DEPENDENT VARIABLE) SUBJECT TO ERROR

Here we desire to fit a second-degree curve, or parabola, to the observed data—i.e., we assume that the functional relation between the dependent variable z and the independent variable x is of the exact form of a parabola:

$$z = \alpha + \beta x + \gamma x^2. \tag{6-89}$$

Again, we postulate that the independent variable x is "free of error", whereas the dependent variable z is measured or obtained with error. Thus the model considered for the observed values x_i and z_i is

$$x_i = u_i \text{ (free of error)} \tag{6-90}$$

$$z_i = \alpha + \beta x_i + \gamma x_i^2 + e_i \text{ (contains error).} \tag{6-91}$$

We will fit the parabola

$$z = a + bx + cx^2 \tag{6-92}$$

to the observed data by determining a, b, and c (which will be estimates of α, β, and γ, respectively) in such a way that the SS of the deviations of the observed values from the fitted values will be a minimum, i.e., by the method of least squares. Actually, we do not have to go through the procedure of finding a, b, and c so that

$$\phi = \sum_{i=1}^{n} (z_i - a - bx - cx^2)^2 \text{ is a minimum}$$

since the method of least squares is very general and we can, as a matter of fact, replace y in Eq. 6-74 for the plane by x^2. Thus we have in a straightforward manner that the coefficients are

$$b = \frac{A_{xz}A_{x^2x^2} - A_{x^2z}A_{xx^2}}{A_{xx}A_{x^2x^2} - A_{xx^2}^2} \tag{6-93}$$

$$c = \frac{A_{xx}A_{x^2z} - A_{xx^2}A_{xz}}{A_{xx}A_{x^2x^2} - A_{xx^2}^2} \tag{6-94}$$

Then the intercept a is found from

$$a = \bar{z} - b\bar{x} - c\overline{x^2} = \frac{1}{n}(\Sigma z_i - b\Sigma x_i - c\Sigma x_i^2) \tag{6-95}$$

where $\overline{x^2}$ denotes the average value of the x^2 observations. The variance of residuals is calculated as the quantity

$$S^2 = \text{est}\sigma_e^2 = \frac{1}{n(n-3)}(A_{zz} - bA_{xz} - cA_{x^2z}). \tag{6-96}$$

The variances of the calculated intercept and coefficients are determined from

$$\text{est}\sigma_a^2 = \frac{nS^2[\Sigma x^2 \Sigma x^4 - (\Sigma x^3)^2]}{A_{xx}A_{x^2x^2} - A_{xx^2}^2} \tag{6-97}$$

$$\text{est}\sigma_b^2 = \frac{nS^2A_{x^2x^2}}{A_{xx}A_{x^2x^2} - A_{xx^2}^2} \tag{6-98}$$

$$\text{est}\sigma_c^2 = \frac{nS^2A_{xx}}{A_{xx}A_{x^2x^2} - A_{xx^2}^2} . \tag{6-99}$$

The variance of a value of z predicted from Eq. 6-92 is given by

$$\sigma_z^2 = \frac{\sigma_e^2}{n} + (x - \bar{x})^2\sigma_b^2 + (x^2 - \overline{x^2})^2\sigma_c^2 + 2(x - \bar{x})(x^2 - \overline{x^2})\sigma_{bc} . \tag{6-100}$$

Estimates of σ_a^2, σ_b^2, and σ_c^2 are therefore given by Eqs. 6-97, 6-98, and 6-99, respectively, whereas an estimate of σ_{bc} is given by

$$\text{est}\sigma_{bc} = \frac{-nA_{xx^2}S^2}{A_{xx}A_{x^2x^2} - A_{xx^2}^2} \tag{6-101}$$

Example 6-3:

A test was conducted* to determine the effect of barrel length on muzzle velocity (MV) for a cal .22 long rifle (Model 37 Remington). The observed data are given in Table 6-5 and each average MV is based on 10 rounds.

*by W.O.L.F. Moore—See APG Firing Record Misc. 017.

TABLE 6-5

RIFLE BARREL LENGTH vs AVERAGE MUZZLE VELOCITY

Barrel Length x, in.	Average Velocity z, ft/s
28	1084
26	1075
24	1091
22	1096
20	1100
18	1098
16	1085
14	1088
12	1085
10	1079
8	1067
6	1040

For the pertinent calculations we find:

$$n = 12 \qquad A_{xx} = 6864$$
$$\Sigma x = 204 \qquad A_{zz} = 35{,}528$$
$$\Sigma z = 12{,}988 \qquad A_{xz} = 8928$$
$$\Sigma x^2 = 4040 \qquad A_{x^2x^2} = 8{,}191{,}040$$
$$\Sigma x^3 = 88{,}128 \qquad A_{x^2z} = 233{,}248$$
$$\Sigma x^4 = 2{,}042{,}720 \qquad A_{xx^2} = 233{,}376$$
$$\Sigma xz = 221{,}540$$
$$\Sigma x^2 z = 4{,}392{,}064$$
$$\Sigma z^2 = 14{,}060{,}306$$

Using Eqs. 6-93 through 6-99, we find

$$b = 10.6286, \qquad c = -0.27435, \qquad a = 994.0115, \qquad S^2 = 42.8464$$
$$\hat{\sigma}_b = 1.547, \qquad \hat{\sigma}_c = 0.0448, \qquad \hat{\sigma}_a = 11.920.$$

Hence

$$t_b = \frac{b}{\hat{\sigma}_b} = 6.87$$

$$t_c = \frac{|c|}{\hat{\sigma}_c} = 6.12$$

$$t_a = \frac{a}{\hat{\sigma}_a} = 83.39.$$

Since a, b, and c are significant at the 0.01 level, we adopt the equation

$$\mathrm{MV} = 994.01 + 10.629(\mathrm{BL}) - 0.2744(\mathrm{BL})^2, \text{ ft/s.}$$

where

BL = barrel length, in.

Since it may be desirable to make linear transformations on the original variables (to reduce effectively the size of numbers in the calculations), the pertinent equations that follow may be of value. Suppose we change the original variables x and z as follows:

$$u_i = c(x_i - h), \qquad v_i = d(z_i - k) \tag{6-102}$$

where c, d, h, and k are constants. Then it can be shown that

$$A_{xx^2} = \left(\frac{1}{c^3}\right) A_{uu^2} + \left(\frac{2h}{c^2}\right) A_{uu} \tag{6-103}$$

$$A_{x^2x^2} = \left(\frac{1}{c^4}\right) A_{u^2u^2} + \left(\frac{4h}{c^3}\right) A_{uu^2} + \left(\frac{4h^2}{c^2}\right) A_{uu} \tag{6-104}$$

$$A_{x^2z} = \left(\frac{1}{c^2 d}\right) A_{u^2v} + \left(\frac{2h}{cd}\right) A_{uv}. \tag{6-105}$$

We had previously shown in Eq. 6-34 that

$$A_{xx} = \left(\frac{1}{c^2}\right) A_{uu}.$$

6-9 THE REGRESSION OF A DEPENDENT VARIABLE (SUBJECT TO ERROR) ON THREE INDEPENDENT VARIABLES (FREE OF ERROR)

For the regression of a dependent variable z containing error on three independent variables—x, y, and u—free of error, we use the model

$$z_i = \alpha + \beta(x_i - \bar{x}) + \gamma(y_i - \bar{y}) + \delta(u_i - \bar{u}) + e_i \tag{6-106}$$

where

δ = true unknown coefficient.

We will estimate z from the equation

$$z = a + b(x - \bar{x}) + c(y - \bar{y}) + d(u - \bar{u}) = (a - b\bar{x} - c\bar{y} - d\bar{u}) + bx + cy + du \tag{6-107}$$

where a, b, c, and d are to be determined by the method of least squares.

In par. 6-8 we extended the model for a plane type of fit in par. 6-7 to that of a quadratic adjustment by simply substituting the square of x, i.e., x^2, for the new variable y, which was added to the previous linear fit of par. 6-2 to obtain the plane. Hence the rather general and useful form of least squares procedures for applications was indicated. Moreover, any number of new or independent variables may be added to the basic line, or the plane, to obtain an extended model with any new variables desired if they seem to give a better or more physically meaningful fit to the original data. However, continuing to add terms to the regression equation obviously will bring up the question of just where to stop with a useful and "best" fit of the data. Moreover, if one continues to add terms, he will, of course, run out of basic data; eventually, he might reach

the point at which all of the parameters cannot be estimated from the least squares procedure. We will briefly discuss in the sequel what is an appropriate number of terms. Moreover, if there is only one independent variable and we are fitting a line, or a quadratic, or a cubic, etc., the use of orthogonal polynomials fits nicely into the use of statistical tests of significance for stopping rules.
If we let

$$\Delta = \frac{1}{n^2}\begin{vmatrix} A_{xx} & A_{xy} & A_{xu} \\ A_{yx} & A_{yy} & A_{yu} \\ A_{ux} & A_{uy} & A_{uu} \end{vmatrix} = \frac{\Delta_1}{n^2} \tag{6-108}$$

say, then from the method of least squares, we find straightforwardly that

$$a = \frac{1}{n}\Sigma z_i. \tag{6-109}$$

The constant term of Eq. 6-107 is $\bar{z} - b\bar{x} - c\bar{y} - d\bar{u}$.
The coefficients b, c, and d are determined from

$$b = \frac{1}{\Delta_1}\begin{vmatrix} A_{xz} & A_{xy} & A_{xu} \\ A_{yz} & A_{yy} & A_{yu} \\ A_{uz} & A_{uy} & A_{uu} \end{vmatrix} \tag{6-110}$$

$$c = \frac{1}{\Delta_1}\begin{vmatrix} A_{xx} & A_{xz} & A_{xu} \\ A_{yx} & A_{yz} & A_{yu} \\ A_{ux} & A_{uz} & A_{uu} \end{vmatrix} \tag{6-111}$$

$$d = \frac{1}{\Delta_1}\begin{vmatrix} A_{xx} & A_{xy} & A_{xz} \\ A_{yx} & A_{yy} & A_{yz} \\ A_{ux} & A_{uy} & A_{uz} \end{vmatrix} \tag{6-112}$$

The variance of residuals is found from

$$S^2 = \frac{1}{n\,(n-4)}\,(A_{zz} - bA_{xz} - cA_{yz} - dA_{uz}). \tag{6-113}$$

The estimated variance of a is

$$\text{est}\sigma_a^2 = \frac{S^2}{n}\,. \tag{6-114}$$

The estimated variances of the coefficients b, c, and d are determined from

$$\text{est}\sigma_b^2 = [nS^2(A_{yy}A_{uu} - A_{yu}^2)]/\Delta_1 \tag{6-115}$$

$$\text{est}\sigma_c^2 = [nS^2(A_{xx}A_{uu} - A_{xu}^2)]/\Delta_1 \tag{6-116}$$

$$\text{est}\sigma_d^2 = [nS^2(A_{xx}A_{yy} - A_{xy}^2)]/\Delta_1\,. \tag{6-117}$$

Eqs. 6-106 through 6-117 give the needed computational forms to fit the linear regression of the dependent variable z on the three independent variables x, y, and u and to make t tests.

Note that if we wanted to fit the cubic

$$z = a + b(x - \bar{x}) + c(x^2 - \overline{x^2}) + d(x^3 - \overline{x^3}) \tag{6-118}$$

where

$\overline{x^3} = \Sigma x^3/n$ = mean of x^3

we could simply replace y_i and u_i in Eq. 6-106 by x_i^2 and x_i^3, respectively.

6-10 FITTING OF ORTHOGONAL POLYNOMIALS FOR THE CASE IN WHICH OBSERVED VALUES OF THE INDEPENDENT VARIABLE ARE AT EQUALLY SPACED INTERVALS

As mentioned in par. 6-9, if we are interested in the regression of a dependent variable on a single independent variable which is observed at equally spaced intervals, the fitting of polynomials can be made with much facility. Thus if we are interested in fitting a polynomial of the form

$$z = a_0 + a_1 x + a_2 x^2 + \cdots + a_r x^r \tag{6-119}$$

for the relation between the variables z and x, and the independent variable x is equally spaced, i.e.,

$$x_i = e + (i - 1)f; \quad i = 1,2,\ldots,n \tag{6-120}$$

then the computations for a least-square fit can be simplified considerably by the use of orthogonal polynomials. Following Fisher and Yates (Ref. 17), we consider polynomials defined as follows:

$$P_r(t_i) = b_0 + b_1 t_i + b_2 t_i^2 + \cdots + b_r t_i^r \tag{6-121}$$

where $i = 1,2,\ldots,n$ represents the number of points; r is the degree of the polynomial ($r = 0,1,2,\ldots$); and the b's are fitted constants to be determined. The variable t_i will be a linear transformation or function of the observed values of the independent variables x_i, which are equally spaced (free of error). Polynomials of the form of Eq. 6-121 are called orthogonal if

$$\sum_{i=1}^{n} P_r(t_i) P_s(t_i) = 0 \text{ for } r \neq s. \tag{6-122}$$

Our procedure will be to fit

$$z_i = A_0 P_0(t_i) + A_1 P_1(t_i) + A_2 P_2(t_i) + \cdots + A_r P_r(t_i) \tag{6-123}$$

by the method of least squares. Hence we determine the coefficients A_0, A_1, etc., so that

$$\phi = \sum_{i=1}^{n} [z_i - A_0 P_0(t_i) - A_1 P_1(t_i) - \cdots - A_r P_r(t_i)]^2 \tag{6-124}$$

is a minimum.

Differentiating Eq. 6-124 with respect to $A_0, A_1, \ldots, A_r$ and setting the derivatives equal to zero, we find the normal equations:

$$\begin{aligned}
A_0 \sum_{i=1}^{n} P_0^2(t_i) + A_1 \sum_{i=1}^{n} P_0(t_i) P_1(t_i) + \cdots + A_r \sum_{i=1}^{n} P_0(t_i) P_r(t_i) &= \sum_{i=1}^{n} P_0(t_i) z_i \\
A_0 \sum_{i=1}^{n} P_0(t_i) P_1(t_i) + A_1 \sum_{i=1}^{n} P_1^2(t_i) + \cdots + A_r \sum_{i=1}^{n} P_1(t_i) P_r(t_i) &= \sum_{i=1}^{n} P_1(t_i) z_i
\end{aligned} \tag{6-125}$$

$$\vdots$$

$$A_0 \sum_{i=1}^{n} P_0(t_i)P_r(t_i) + A_1 \sum_{i=1}^{n} P_1(t_i)P_r(t_i) + \cdots + A_r \sum_{i=1}^{n} P_r^2(t_i) = \sum_{i=1}^{n} P_r(t_i)z_i.$$

Note that the cross-product terms not on the principal diagonal are of the type

$$\sum_{i=1}^{n} P_r(t_i)P_s(t_i), \text{ where we have that } r \neq s.$$

But these cross-product polynomials for which $r \neq s$ are zero if the polynomials are orthogonal. Thus for orthogonal polynomials we have solutions immediately for the A's, which are

$$A_0 = \frac{\sum_{i=1}^{n} P_0(t_i)z_i}{\sum_{i=1}^{n} P_0^2(t_i)} \tag{6-126}$$

$$A_1 = \frac{\sum_{i=1}^{n} P_1(t_i)z_i}{\sum_{i=1}^{n} P_1^2(t_i)} \tag{6-127}$$

$$\vdots$$

$$A_r = \frac{\sum_{i=1}^{n} P_r(t_i)z_i}{\sum_{i=1}^{n} P_r^2(t_i)}. \tag{6-128}$$

The problem then is to find the polynomials $P_r(t_i)$ that result in orthogonality. This can be done if we put

$$t_i = (x_i - \bar{x})/f \tag{6-129}$$

(where f is the width of the interval between the observations x_i) and choose the $P_r(t_i)$ as follows:

$$\left.\begin{aligned}
P_0(t_i) &= 1 = \xi_0' \quad \text{(in Table 6-6 taken from Fisher and Yates, Ref. 17)} \\
P_1(t_i) &= \lambda_1 t_i = \xi_1' \\
P_2(t_i) &= \lambda_2\left[t_i^2 - \left(\frac{n^2-1}{12}\right)\right] = \xi_2' \\
P_3(t_i) &= \lambda_3\left[t_i^3 - \left(\frac{3n^2-7}{20}\right)t_i\right] = \xi_3' \\
P_4(t_i) &= \lambda_4\left[t_i^4 - \left(\frac{3n^2-13}{14}\right)t_i^2 + \frac{3(n^2-1)(n^2-9)}{560}\right] = \xi_4'
\end{aligned}\right\} \tag{6-130}$$

etc.

The λ_i's are constants that depend on the number of points n and are chosen so that for values of t_i (which are positive or negative integers or 0), the polynomials in the brackets of Eq. 6-130 are whole numbers. The general recurrence equation for the $P_r(t_i)$ or ξ_r' is

$$\xi_{r+1} = \xi_1 \xi_r - \frac{r^2(n^2 - r^2)}{4(4r^2 - 1)} \xi_{r-1}, \quad r = 1,2, \ldots \tag{6-131}$$

where

$$\xi'_r = \xi_r \xi_r.$$

Fisher and Yates' (Ref. 17) Table XXIII, entitled "Orthogonal Polynomials", pp. 62-8, gives the required values of the orthogonal polynomials $P_r(t)$, or ξ'_r, for $r = 1,2, \ldots,5$ (i.e., through the fifth degree) and for the number of points n up through $n = 52$. We reproduce, with permission, Fisher and Yates' Table XXIII as Table 6-6. The values of ξ'_2 and ξ'_4 are symmetrical about their middle values, and the ξ'_1, ξ'_3, and ξ'_5 are also symmetrical except that the values in the first half of each sequence are the negatives of those in the last half. For this reason, only half of the values (i.e., the upper ones) are tabulated for $n \geq 9$. The first two rows under each table give values of the sum of the squares of the ξ'_r, and the third or last row just below each table gives values of the λ_r.

It can be shown that an ordinary polynomial

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k \tag{6-132}$$

can always be expressed in terms of orthogonal polynomials for any specified set of values of x. For example, when $x = 1,2,3, \ldots,7$

$$y = -35 + 59x - 21x^2 + 2x^3$$

can be written in the form

$$y = 5 + (-4 + x) + 3(12 - 8x + x^2) + 12(-6 + \frac{41}{6} x - 2x^2 + \frac{1}{6} x^3)$$

where the polynomials in parentheses are orthogonal, as seen in Table 6-7.

Table 6-7, therefore, exhibits the required properties of the orthogonal polynomials.

Example 6-4:

Using the data of Example 6-3 for length of barrel of the cal .22 long rifle versus the average muzzle velocity, we arrange the computations as in Table 6-8, where the values of ξ'_r are taken from Table 6-6.

Calculations follow with $\bar{x} = 17$, $n = 12$, and the data from Table 6-6:

$$t_i = \xi_1 = (x_i - \bar{x})/f = (x_i - 17)/2; \quad \xi'_1 = \lambda_1 t_i = 2t_i$$

$$\xi'_2 = \lambda_2[t_i^2 - (n^2 - 1)/12] = 3(t_i^2 - 143/12)$$

etc., as in Eq. 6-130.

The mean velocity from Table 6-8 data is

$$\bar{z} = 2183 + 2188 + 2181 + 2170 + 2142 + 2124)/12 = 1082.33$$

$$z = a + b\xi'_1 + c\xi'_2 + d\xi'_3, \quad \text{where } a = \bar{z} = 1082.33$$

$$b = \Sigma \xi'_{1_i} d_i / 572 = 744/572 = 1.3007$$

$$c = \Sigma \xi'_{2_i} s_i / 12{,}012 = -4394/12{,}012 = -0.3658$$

$$d = \Sigma \xi'_{3_i} d_i / 5148 = 582/5148 = 0.1131.$$

The analysis of variance (ANOVA) is put in the form of Table 6-9.

TABLE 6-6

TABLES OF ORTHOGONAL POLYNOMIALS (Ref. 17)

(Values from $n = 32$ to $n = 51$ are due to V. Satakopan)

3		4			5				6					7					8				
ξ'_1	ξ'_2	ξ'_1	ξ'_2	ξ'_3	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
−1	+1	−3	+1	−1	−2	+2	−1	+1	−5	+5	−5	+1	−1	−3	+5	−1	+3	−1	−7	+7	−7	+7	−7
0	−2	−1	−1	+3	−1	−1	+2	−4	−3	−1	+7	−3	+5	−2	0	+1	−7	+4	−5	+1	+5	−13	+23
+1	+1	+1	−1	−3	0	−2	0	+6	−1	−4	+4	+2	−10	−1	−3	+1	+1	−5	−3	−3	+7	−3	−17
		+3	+1	+1	+1	−1	−2	−4	+1	−4	−4	+2	+10	0	−4	0	+6	0	−1	−5	+3	+9	−15
					+2	+2	+1	+1	+3	−1	−7	−3	−5	+1	−3	−1	+1	+5	+1	−5	−3	+9	+15
									+5	+5	+5	+1	+1	+2	0	−1	−7	−4	+3	−3	−7	−3	+17
														+3	+5	+1	+3	+1	+5	+1	−5	−13	−23
																			+7	+7	+7	+7	+7
2	6	20	4	20	10	14	10	70	70	84	180	28	252	28	84	6	154	84	168	168	264	616	2184
1	3	2	1	$\frac{10}{3}$	1	1	$\frac{5}{6}$	$\frac{35}{12}$	2	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{7}{12}$	$\frac{21}{10}$	1	1	$\frac{1}{6}$	$\frac{7}{12}$	$\frac{7}{20}$	2	1	$\frac{2}{3}$	$\frac{7}{12}$	$\frac{7}{10}$

9					10					11					12				
ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−20	0	+18	0	+1	−4	−12	+18	+6	0	−10	0	+6	0	+1	−35	−7	+28	+20
+1	−17	−9	+9	+9	+3	−3	−31	+3	+11	+1	−9	−14	+4	+4	+3	−29	−19	+12	+44
+2	−8	−13	−11	+4	+5	−1	−35	−17	+1	+2	−6	−23	−1	+4	+5	−17	−25	−13	+39
+3	+7	−7	−21	−11	+7	+2	−14	−22	−14	+3	−1	−22	−6	−1	+7	+1	−21	−33	−21
+4	+28	+14	+14	+4	+9	+6	+42	+18	+6	+4	+6	−6	−6	−6	+9	+25	−3	−27	−57
										+5	+15	+30	+6	+3	+11	+55	+33	+33	+33
60	2,772	990	2,002	468	330	132	8,580	2,860	780	110	858	4,290	286	156	572	12,012	5,148	8,008	15,912
1	3	$\frac{5}{6}$	$\frac{7}{12}$	$\frac{3}{20}$	2	$\frac{1}{2}$	$\frac{5}{3}$	$\frac{5}{12}$	$\frac{1}{10}$	1	1	$\frac{5}{6}$	$\frac{1}{12}$	$\frac{1}{40}$	2	3	$\frac{2}{3}$	$\frac{7}{24}$	$\frac{3}{20}$

13					14					15				
ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−14	0	+84	0	+1	−8	−24	+108	+60	0	−56	0	+756	0
+1	−13	−4	+64	+20	+3	−7	−67	+63	+145	+1	−53	−27	+621	+675
+2	−10	−7	+11	+26	+5	−5	−95	−13	+139	+2	−44	−49	+251	+1000
+3	−5	−8	−54	+11	+7	−2	−98	−92	+28	+3	−29	−61	−249	+751
+4	+2	−6	−96	−18	+9	+2	−66	−132	−132	+4	−8	−58	−704	−44
+5	+11	0	−66	−33	+11	+7	+11	−77	−187	+5	+19	−35	−869	−979
+6	+22	+11	+99	+22	+13	+13	+143	+143	+143	+6	+52	+13	−429	−1144
										+7	+91	+91	+1001	+1001
182	2,002	572	68,068	6,188	910	728	97,240	136,136	235,144	280	37,128	39,780	6,466,460	10,581,480
1	1	$\frac{1}{6}$	$\frac{7}{12}$	$\frac{7}{120}$	2	$\frac{1}{2}$	$\frac{5}{3}$	$\frac{7}{12}$	$\frac{7}{30}$	1	3	$\frac{5}{6}$	$\frac{35}{12}$	$\frac{21}{20}$

16					17					18				
ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5	ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
+1	−21	−63	+189	+45	0	−24	0	+36	0	+1	−40	−8	+44	+220
+3	−19	−179	+129	+115	+1	−23	−7	+31	+55	+3	−37	−23	+33	+583
+5	−15	−265	+23	+131	+2	−20	−13	+17	+88	+5	−31	−35	+13	+733
+7	−9	−301	−101	+77	+3	−15	−17	−3	+83	+7	−22	−42	−12	+588
+9	−1	−267	−201	−33	+4	−8	−18	−24	+36	+9	−10	−42	−36	+156
+11	+9	−143	−221	−143	+5	+1	−15	−39	−39	+11	+5	−33	−51	−429
+13	+21	+91	−91	−143	+6	+12	−7	−39	−104	+13	+23	−13	−47	−871
+15	+35	+455	+273	+143	+7	+25	+7	−13	−91	+15	+44	+20	−12	−676
					+8	+40	+28	+52	+104	+17	+68	+68	+68	+884
1,360	5,712	1,007,760	470,288	201,552	408	7,752	3,876	16,796	100,776	1,938	23,256	23,256	28,424	6,953,544
2	1	$\frac{10}{3}$	$\frac{7}{12}$	$\frac{1}{10}$	1	1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{20}$	2	$\frac{3}{2}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{3}{10}$

(cont'd on next page)

TABLE 6-6 (Cont'd)

19

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−30	0	+396	0
+1	−29	−44	+352	+44
+2	−26	−83	+227	+74
+3	−21	−112	+42	+79
+4	−14	−126	−168	+54
+5	−5	−120	−354	+3
+6	+6	−89	−453	−58
+7	+19	−28	−388	−98
+8	+34	+68	−68	−68
+9	+51	+204	+612	+102
570		213,180		89,148
	13,566		2,288,132	
1	1	$\frac{5}{6}$	$\frac{7}{12}$	$\frac{1}{40}$

20

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
+1	−33	−99	+1188	+396
+3	−31	−287	+948	+1076
+5	−27	−445	+503	+1441
+7	−21	−553	−77	+1351
+9	−13	−591	−687	+771
+11	−3	−539	−1187	−187
+13	+9	−377	−1402	−1222
+15	+23	−85	−1122	−1802
+17	+39	+357	−102	−1122
+19	+57	+969	+1938	+1938
2,660		4,903,140		31,201,800
	17,556		22,881,320	
2	1	$\frac{10}{8}$	$\frac{35}{24}$	$\frac{7}{20}$

21

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−110	0	+594	0
+1	−107	−54	+540	+1404
+2	−98	−103	+385	+2444
+3	−83	−142	+150	+2819
+4	−62	−166	−130	+2354
+5	−35	−170	−406	+1063
+6	−2	−149	−615	−788
+7	+37	−98	−680	−2618
+8	+82	−12	−510	−3468
+9	+133	+114	0	−1938
+10	+190	+285	+969	+3876
770		432,630		121,687,020
	201,894		5,720,330	
1	3	$\frac{5}{6}$	$\frac{7}{12}$	$\frac{21}{40}$

22

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
+1	−20	−12	+702	+390
+3	−19	−35	+585	+1079
+5	−17	−55	+365	+1509
+7	−14	−70	+70	+1554
+9	−10	−78	−258	+1158
+11	−5	−77	−563	+363
+13	+1	−65	−775	−663
+15	+8	−40	−810	−1598
+17	+16	0	−570	−1938
+19	+25	+57	+57	−969
+21	+35	+133	+1197	+2261
3,542		96,140		40,562,340
	7,084		8,748,740	
2	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{7}{12}$	$\frac{7}{30}$

23

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−44	0	+858	0
+1	−43	−13	+793	+65
+2	−40	−25	+605	+116
+3	−35	−35	+315	+141
+4	−28	−42	−42	+132
+5	−19	−45	−417	+87
+6	−8	−43	−747	+12
+7	+5	−35	−955	−77
+8	+20	−20	−950	−152
+9	+37	+3	−627	−171
+10	+56	+35	+133	−76
+11	+77	+77	+1463	+209
1,012		32,890		340,8[illegible]
	35,420		13,123,110	
1	1	$\frac{1}{6}$	$\frac{7}{12}$	$\frac{1}{60}$

24

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
+1	−143	−143	+143	+715
+3	−137	−419	+123	+2005
+5	−125	−665	+85	+2893
+7	−107	−861	+33	+3171
+9	−83	−987	−27	+2721
+11	−53	−1023	−87	+1551
+13	−17	−949	−137	−169
+15	+25	−745	−165	−2071
+17	+73	−391	−157	−3553
+19	+127	+133	−97	−3743
+21	+187	+847	+33	−1463
+23	+253	+1771	+253	+4807
4,600		17,760,600		177,928,920
	394,680		394,680	
2	3	$\frac{10}{3}$	$\frac{7}{12}$	$\frac{3}{10}$

25

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−52	0	+858	0
+1	−51	−77	+803	+275
+2	−48	−149	+643	+500
+3	−43	−211	+393	+631
+4	−36	−258	+78	+636
+5	−27	−285	−267	+501
+6	−16	−287	−597	+236
+7	−3	−259	−857	−119
+8	+12	−196	−982	−488
+9	+29	−93	−897	−753
+10	+48	+55	−517	−748
+11	+69	+253	+253	−253
+12	+92	+506	+1518	+1012
1,300		1,480,050		7,803,900
	53,820		14,307,150	
1	1	$\frac{5}{6}$	$\frac{5}{12}$	$\frac{1}{20}$

26

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
+1	−28	−84	+1386	+330
+3	−27	−247	+1221	+935
+5	−25	−395	+905	+1381
+7	−22	−518	+466	+1582
+9	−18	−606	−54	+1482
+11	−13	−649	−599	+1067
+13	−7	−637	−1099	+377
+15	0	−560	−1470	−482
+17	+8	−408	−1614	−1326
+19	+17	−171	−1419	−1881
+21	+27	+161	−759	−1771
+23	+38	+598	+506	−506
+25	+50	+1150	+2530	+2530
5,850		7,803,900		48,384,180
	16,380		40,060,020	
2	$\frac{1}{2}$	$\frac{5}{3}$	$\frac{7}{12}$	$\frac{1}{10}$

27

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−182	0	+1638	0
+1	−179	−18	+1548	+3960
+2	−170	−35	+1285	+7304
+3	−155	−50	+870	+9479
+4	−134	−62	+338	+10058
+5	−107	−70	−262	+8803
+6	−74	−73	−867	+5728
+7	−35	−70	−1400	+1162
+8	+10	−60	−1770	−4188
+9	+61	−42	−1872	−9174
+10	+118	−15	−1587	−12144
+11	+181	+22	−782	−10879
+12	+250	+70	+690	−2530
+13	+325	+130	+2990	+16445
1,638		101,790		2,032,135,560
	712,530		56,448,210	
1	3	$\frac{1}{6}$	$\frac{7}{12}$	$\frac{21}{40}$

(cont'd on next page)

TABLE 6-6 (Cont'd)

28

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
+1	−65	−39	+936	+1560
+3	−63	−115	+840	+4456
+5	−59	−185	+655	+6701
+7	−53	−245	+395	+7931
+9	−45	−291	+81	+7887
+11	−35	−319	−259	+6457
+13	−23	−325	−590	+3718
+15	−9	−305	−870	−22
+17	+7	−255	−1050	−4182
+19	+25	−171	−1074	−7866
+21	+45	−49	−879	−9821
+23	+67	+115	−395	−8395
+25	+91	+325	+455	−1495
+27	+117	+585	+1755	+13455

7,398 | 95,004 | 2,103,660 | 19,634,160 | 1,354,757,040

2 | 1 | $\frac{2}{3}$ | $\frac{7}{24}$ | $\frac{7}{20}$

29

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−70	0	+2184	0
+1	−69	−104	+2080	+1768
+2	−66	−203	+1775	+3298
+3	−61	−292	+1290	+4373
+4	−54	−366	+660	+4818
+5	−45	−420	−66	+4521
+6	−34	−449	−825	+3454
+7	−21	−448	−1540	+1694
+8	−6	−412	−2120	−556
+9	+11	−336	−2460	−2946
+10	+30	−215	−2441	−4958
+11	+51	−44	−1930	−5885
+12	+74	+182	−780	−4810
+13	+99	+468	+1170	−585
+14	+126	+819	+4095	+8190

2,030 | 113,274 | 4,207,320 | 107,987,880 | 500,671,080

1 | 1 | $\frac{5}{6}$ | $\frac{7}{12}$ | $\frac{7}{40}$

30

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
+1	−112	−112	+12376	+1768
+3	−109	−331	+11271	+5083
+5	−103	−535	+9131	+7753
+7	−94	−714	+6096	+9408
+9	−82	−858	+2376	+9768
+11	−67	−957	−1749	+8679
+13	−49	−1001	−5929	+6149
+15	−28	−980	−9744	+2384
+17	−4	−884	−12704	−2176
+19	+23	−703	−14249	−6821
+21	+53	−427	−13749	−10535
+23	+86	−46	−10504	−11960
+25	+122	+450	−3744	−9360
+27	+161	+1071	+7371	−585
+29	+203	+1827	+23751	+16965

8,990 | 302,064 | 21,360,240 | 3,671,587,920 | 2,145,733,200

2 | $\frac{3}{2}$ | $\frac{5}{3}$ | $\frac{35}{12}$ | $\frac{3}{10}$

31

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
+0	−80	0	+408	0
+1	−79	−119	+391	+221
+2	−76	−233	+341	+416
+3	−71	−337	+261	+561
+4	−64	−426	+156	+636
+5	−55	−495	+33	+627
+6	−44	−539	−99	+528
+7	−31	−553	−229	+343
+8	−16	−532	−344	+88
+9	+1	−471	−429	−207
+10	+20	−365	−467	−496
+11	+41	−209	−439	−715
+12	+64	+2	−324	−780
+13	+89	+273	−99	−585
+14	+116	+609	+261	0
+15	+145	+1015	+783	+1131

2,480 | 158,224 | 6,724,520 | 4,034,712 | 9,536,592

1 | 1 | $\frac{5}{6}$ | $\frac{1}{12}$ | $\frac{1}{60}$

32

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
+1	−85	−51	+459	+255
+3	−83	−151	+423	+737
+5	−79	−245	+353	+1137
+7	−73	−329	+253	+1407
+9	−65	−399	+129	+1509
+11	−55	−451	−11	+1419
+13	−43	−481	−157	+1131
+15	−29	−485	−297	+661
+17	−13	−459	−417	+51
+19	+5	−399	−501	−627
+21	+25	−301	−531	−1267
+23	+47	−161	−487	−1725
+25	+71	+25	−347	−1815
+27	+97	+261	−87	−1305
+29	+125	+551	+319	+87
+31	+155	+899	+899	+2697

10,912 | 185,504 | 5,379,616 | 5,379,616 | 54,285,216

2 | 1 | $\frac{2}{3}$ | $\frac{1}{12}$ | $\frac{1}{30}$

33

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−272	0	+3672	0
+1	−269	−27	+3537	+2565
+2	−260	−53	+3139	+4864
+3	−245	−77	+2499	+6649
+4	−224	−98	+1652	+7708
+5	−197	−115	+647	+7883
+6	−164	−117	−453	+7088
+7	−125	−133	−1571	+5327
+8	−80	−132	−2616	+2712
+9	−29	−123	−3483	−519
+10	+28	−105	−4053	−3984
+11	+91	−77	−4193	−7139
+12	+160	−38	−3756	−9260
+13	+235	+13	−2581	−9425
+14	+316	+77	−493	−6496
+15	+403	+155	+2697	+899
+16	+496	+248	+7192	+14384

2,992 | 1,947,792 | 417,384 | 348,330,136 | 1,547,128,656

1 | 3 | $\frac{1}{6}$ | $\frac{7}{12}$ | $\frac{3}{20}$

(cont'd on next page)

TABLE 6-6 (Cont'd)

34

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
1	−48	−144	+4104	+6840
3	−47	−427	+3819	+19855
5	−45	−695	+3263	+30917
7	−42	−938	+2464	+38864
9	−38	−1146	+1464	+42744
11	−33	−1309	+319	+41899
13	−27	−1417	−901	+36049
15	−20	−1460	−2112	+25376
17	−12	−1428	−3216	+10608
19	−3	−1311	−4101	−6897
21	+7	−1099	−4641	−25067
23	+18	−782	−4696	−41032
25	+30	−350	−4112	−51040
27	+43	+207	−2721	−50373
29	+57	+899	−341	−33263
31	+72	+1736	+3224	+7192
33	+88	+2728	+8184	+79112

13,090 51,477,360 46,929,569,232
62,832 456,432,592
2 $\frac{1}{2}$ $\frac{5}{3}$ $\frac{7}{12}$ $\frac{7}{10}$

35

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−102	0	+23256	0
1	−101	−152	+22496	+3800
2	−98	−299	+20251	+7250
3	−93	−436	+16626	+10021
4	−86	−558	+11796	+11826
5	−77	−660	+6006	+12441
6	−66	−737	−429	+11726
7	−53	−784	−7124	+9646
8	−38	−796	−13624	+6292
9	−21	−768	−19404	+1902
10	−2	−695	−23869	−3118
11	+19	−572	−26354	−8173
12	+42	−394	−26124	−12458
13	+67	−156	−22374	−14937
14	+94	+147	−14229	−14322
15	+123	+520	−744	−9052
16	+154	+968	+19096	+2728
17	+187	+1496	+46376	+23188

3,570 15,775,320 4,045,652,520
290,598 14,834,059,240
1 1 $\frac{5}{6}$ $\frac{35}{12}$ $\frac{7}{40}$

36

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
1	−323	−323	+2584	+12920
3	−317	−959	+2424	+37640
5	−305	−1565	+2111	+59063
7	−287	−2121	+1659	+75201
9	−263	−2607	+1089	+84381
11	−233	−3003	+429	+85371
13	−197	−3289	−286	+77506
15	−155	−3445	−1014	+60814
17	−107	−3451	−1706	+36142
19	−53	−3287	−2306	+5282
21	+7	−2933	−2751	−28903
23	+73	−2369	−2971	−62353
25	+145	−1575	−2889	−89685
27	+223	−531	−2421	−104067
29	+307	+783	−1476	−97092
31	+397	+2387	+44	−58652
33	+493	+4301	+2244	+23188
35	+595	+6545	+5236	+162316

15,540 307,618,740 199,046,103,984
3,011,652 191,407,216
2 3 $\frac{10}{3}$ $\frac{7}{24}$ $\frac{21}{20}$

37

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−114	0	+5814	0
1	−113	−34	+5644	+680
2	−110	−67	+5141	+1304
3	−105	−98	+4326	+1819
4	−98	−126	+3234	+2178
5	−89	−150	+1914	+2343
6	−78	−169	+429	+2288
7	−65	−182	−1144	+2002
8	−50	−188	−2714	+1492
9	−33	−186	−4176	+786
10	−14	−175	−5411	−64
11	+7	−154	−6286	−979
12	+30	−122	−6654	−1850
13	+55	−78	−6354	−2535
14	+82	−21	−5211	−2856
15	+111	+50	−3036	−2596
16	+142	+136	+374	−1496
17	+175	+238	+5236	+748
18	+210	+357	+11781	+4488

4,218 932,178 152,877,192
383,838 980,961,982
1 1 $\frac{1}{6}$ $\frac{7}{12}$ $\frac{1}{40}$

38

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
1	−60	−36	+918	+1530
3	−59	−107	+867	+4471
5	−57	−175	+767	+7061
7	−54	−238	+622	+9086
9	−50	−294	+438	+10362
11	−45	−341	+223	+10747
13	−39	−377	−13	+10153
15	−32	−400	−258	+8558
17	−24	−408	−498	+6018
19	−15	−399	−717	+2679
21	−5	−371	−897	−1211
23	+6	−322	−1018	−5290
25	+18	−250	−1058	−9070
27	+31	−153	−993	−11925
29	+45	−29	−797	−13079
31	+60	+124	−442	−11594
33	+76	+308	+102	−6358
35	+ 93	+525	+867	+3927
37	+111	+777	+1887	+20757

18,278 4,496,388 3,286,859,628
109,668 25,479,532
2 $\frac{1}{2}$ $\frac{1}{3}$ $\frac{1}{12}$ $\frac{1}{10}$

39

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−380	0	+1026	0
1	−377	−189	+999	+5049
2	−368	−373	+919	+9724
3	−353	−547	+789	+13669
4	−332	−706	+614	+16564
5	−305	−845	+401	+18143
6	−272	−959	+159	+18212
7	−233	−1043	−101	+16667
8	−188	−1092	−366	+13512
9	−137	−1101	−621	+8877
10	−80	−1065	−849	+3036
11	−17	−979	−1031	−3575
12	+52	−838	−1146	−10340
13	+127	−637	−1171	−16445
14	+208	−371	−1081	−20860
15	+295	−35	−849	−22321
16	+388	+376	−446	−19312
17	+487	+867	+159	−10047
18	+592	+1443	+999	+7548
19	+703	+2109	+2109	+35853

4,940 33,722,910 9,860,578,884
4,496,388 32,224,114
1 3 $\frac{5}{6}$ $\frac{3}{12}$ $\frac{3}{20}$

(cont'd on next page)

TABLE 6-6 (Cont'd)

40

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
1	−133	−399	+39501	+627
3	−131	−1187	+37521	+1837
5	−127	−1945	+33631	+2917
7	−121	−2653	+27971	+3787
9	−113	−3291	+20751	+4377
11	−103	−3839	+12251	+4631
13	−91	−4277	+2821	+4511
15	−77	−4585	−7119	+4001
17	−61	−4743	−17079	+3111
19	−43	−4731	−26499	+1881
21	−23	−4529	−34749	+385
23	−1	−4117	−41129	−1265
25	+23	−3475	−44869	−2915
27	+49	−2583	−45129	−4365
29	+77	−1421	−40999	−5365
31	+107	+31	−31499	−5611
33	+139	+1793	−15579	−4741
35	+173	+3885	+7881	−2331
37	+209	+6327	+40071	+2109
39	+247	+9139	+82251	+9139

21,320 644,482,280 644,482,280
567,112 49,625,135,560
2 1 $\frac{10}{3}$ $\frac{35}{12}$ $\frac{1}{30}$

41

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−140	0	+8778	0
1	−139	−209	+8569	+4807
2	−136	−413	+7949	+9292
3	−131	−607	+6939	+13147
4	−124	−786	+5574	+16092
5	−115	−945	+3903	+17889
6	−104	−1079	+1989	+18356
7	−91	−1183	−91	+17381
8	−76	−1252	−2246	+14936
9	−59	−1281	−4371	+11091
10	−40	−1265	−6347	+6028
11	−19	−1199	−8041	+55
12	+4	−1078	−9306	−6380
13	+29	−897	−9981	−12675
14	+56	−651	−9891	−18060
15	+85	−335	−8847	−21583
16	+116	+56	−6646	−22096
17	+149	+527	−3071	−18241
18	+184	+1083	+2109	−8436
19	+221	+1729	+9139	+9139
20	+260	+2470	+18278	+36556

5,740 47,900,710 10,376,164,708
641,732 2,481,256,778
1 1 $\frac{5}{6}$ $\frac{7}{12}$ $\frac{7}{60}$

42

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
1	−220	−44	+9614	+48070
3	−217	−131	+9177	+141151
5	−211	−215	+8317	+225181
7	−202	−294	+7062	+294546
9	−190	−366	+5454	+344262
11	−175	−429	+3549	+370227
13	−157	−481	+1417	+369473
15	−136	−520	−858	+340418
17	−112	−544	−3178	+283118
19	−85	−551	−5431	+199519
21	−55	−539	−7491	+93709
23	−22	−506	−9218	−27830
25	+14	−450	−10458	−155970
27	+53	−369	−11043	−278685
29	+95	−261	−10791	−380799
31	+140	−124	−9506	−443734
33	+188	+44	−6978	−445258
35	+239	+245	−2983	−359233
37	+293	+481	+2717	−155363
39	+350	+754	+10374	+201058
41	+410	+1066	+20254	+749398

24,682 9,075,924 4,389,117,671,484
1,629,012 3,084,805,724
2 $\frac{3}{2}$ $\frac{1}{3}$ $\frac{7}{12}$ $\frac{91}{10}$

43

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−154	0	+10626	0
1	−153	−46	+10396	+8740
2	−150	−91	+9713	+16948
3	−145	−134	+8598	+24113
4	−138	−174	+7086	+29766
5	−129	−210	+5226	+33501
6	−118	−241	+3081	+34996
7	−105	−266	+728	+34034
8	−90	−284	−1742	+30524
9	−73	−294	−4224	+24522
10	−54	−295	−6599	+16252
11	−33	−286	−8734	+6127
12	−10	−266	−10482	−5230
13	+15	−234	−11682	−16965
14	+42	−189	−12159	−27972
15	+71	−130	−11724	−36872
16	+102	−56	−10174	−41992
17	+135	+34	−7292	−41344
18	+170	+141	−2847	−32604
19	+207	+266	+3406	−13091
20	+246	+410	+11726	+20254
21	+287	+574	+22386	+70889

6,622 2,676,234 39,541,600,644
814,506 3,815,417,606
1 1 $\frac{1}{6}$ $\frac{7}{12}$ $\frac{7}{40}$

44

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
1	−161	−483	+5796	+1380
3	−159	−1439	+5556	+4060
5	−155	−2365	+5083	+6503
7	−149	−3241	+4391	+8561
9	−141	−4047	+3501	+10101
11	−131	−4763	+2441	+11011
13	−119	−5369	+1246	+11206
15	−105	−5845	−42	+10634
17	−89	−6171	−1374	+9282
19	−71	−6327	−2694	+7182
21	−51	−6293	−3939	+4417
23	−29	−6049	−5039	+1127
25	−5	−5575	−5917	−2485
27	+21	−4851	−6489	−6147
29	+49	−3857	−6664	−9512
31	+79	−2573	−6344	−12152
33	+111	−979	−5424	−13552
35	+145	+945	−3792	−13104
37	+181	+3219	−1329	−10101
39	+219	+5863	+2091	−3731
41	+259	+8897	+6601	+6929
43	+301	+12341	+12341	+22919

28,380 1,257,829,980 4,162,273,752
913,836 1,173,974,648
2 1 $\frac{10}{3}$ $\frac{7}{24}$ $\frac{1}{20}$

45

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−506	0	+9108	0
1	−503	−252	+8928	+4500
2	−494	−499	+8393	+8750
3	−479	−736	+7518	+12509
4	−458	−958	+6328	+15554
5	−431	−1160	+4858	+17689
6	−398	−1337	+3153	+18754
7	−359	−1484	+1268	+18634
8	−314	−1596	−732	+17268
9	−263	−1668	−2772	+14658
10	−206	−1695	−4767	+10878
11	−143	−1672	−6622	+6083
12	−74	−1594	−8232	+518
13	+1	−1456	−9482	−5473
14	+82	−1253	−10247	−11438
15	+169	−980	−10392	−16808
16	+262	−632	−9772	−20888
17	+361	−204	−8232	−22848
18	+466	+309	−5607	−21714
19	+577	+912	−1722	−16359
20	+694	+1610	+3608	−5494
21	+817	+2408	+10578	+12341
22	+946	+3311	+19393	+38786

7,590 92,036,340 12,006,558,900
9,203,634 2,934,936,620
1 3 $\frac{5}{6}$ $\frac{5}{12}$ $\frac{3}{40}$

(cont'd on next page)

TABLE 6-6 (Cont'd)

46

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
1	−88	−264	+1980	+3300
3	−87	−787	+1905	+9725
5	−85	−1295	+1757	+15631
7	−82	−1778	+1540	+20692
9	−78	−2226	+1260	+24612
11	−73	−2629	+925	+27137
13	−67	−2977	+545	+28067
15	−60	−3260	+132	+27268
17	−52	−3468	−300	+24684
19	−43	−3591	−735	+20349
21	−33	−3619	−1155	+14399
23	−22	−3542	−1540	+7084
25	−10	−3350	−1868	−1220
27	+3	−3033	−2115	−9999
29	+17	−2581	−2255	−18589
31	+32	−1984	−2260	−26164
33	+48	−1232	−2100	−31724
35	+65	−315	−1743	−34083
37	+83	+777	−1155	−31857
39	+102	+2054	−300	−23452
41	+122	+3526	+860	−7052
43	+143	+5203	+2365	+19393
45	+165	+7095	+4257	+58179

32,430 429,502,920 27,214,866,840
285,384 143,167,640
2 $\frac{1}{2}$ $\frac{5}{3}$ $\frac{1}{12}$ $\frac{1}{10}$

47

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−184	0	+15180	0
1	−183	−55	+14905	+3575
2	−180	−109	+14087	+6968
3	−175	−161	+12747	+10003
4	−168	−210	+10920	+12516
5	−159	−255	+8655	+14361
6	−148	−295	+6015	+15416
7	−135	−329	+3077	+15589
8	−120	−356	−68	+14824
9	−103	−375	−3315	+13107
10	−84	−385	−6545	+10472
11	−63	−385	−9625	+7007
12	−40	−374	−12408	+2860
13	−15	−351	−14733	−1755
14	+12	−315	−16425	−6552
15	+41	−265	−17295	−11167
16	+72	−200	−17140	−15152
17	+105	−119	−15743	−17969
18	+140	−21	−12873	−18984
19	+177	+95	−8285	−17461
20	+216	+230	−1720	−12556
21	+257	+385	+7095	−3311
22	+300	+561	+18447	+11352
23	+345	+759	+32637	+32637

8,648 4,994,220 8,629,104,120
1,271,256 8,518,474,580
1 1 $\frac{1}{6}$ $\frac{7}{12}$ $\frac{1}{20}$

48

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
1	−575	−115	+16445	+82225
3	−569	−343	+15873	+242671
5	−557	−565	+14743	+391231
7	−539	−777	+13083	+520401
9	−515	−975	+10935	+623307
11	−485	−1155	+8355	+693957
13	−449	−1313	+5413	+727493
15	−407	−1445	+2193	+720443
17	−359	−1547	−1207	+670973
19	−305	−1615	−4675	+579139
21	−245	−1645	−8085	+447139
23	−179	−1633	−11297	+279565
25	−107	−1575	−14157	+83655
27	−29	−1467	−16497	−130455
29	+55	−1305	−18135	−349479
31	+145	−1085	−18875	−556729
33	+241	−803	−18507	−731863
35	+343	−455	−16807	−850633
37	+451	−37	−13537	−884633
39	+565	+455	−8445	−801047
41	+685	+1025	−1265	−562397
43	+811	+1677	+8283	−126291
45	+943	+2415	+20493	+554829
47	+1081	+3243	+35673	+1533939

36,848 92,620,080 19,208,385,771,120
12,712,560 10,301,411,120
2 3 $\frac{2}{3}$ $\frac{7}{12}$ $\frac{21}{10}$

49

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−200	0	+17940	0
1	−199	−299	+17641	+9867
2	−196	−593	+16751	+19272
3	−191	−877	+15291	+27767
4	−184	−1146	+13296	+34932
5	−175	−1395	+10815	+40389
6	−164	−1619	+7911	+43816
7	−151	−1813	+4661	+44961
8	−136	−1972	+1156	+43656
9	−119	−2091	−2499	+39831
10	−100	−2165	−6185	+33528
11	−79	−2189	−9769	+24915
12	−56	−2158	−13104	+14300
13	−31	−2067	−16029	+2145
14	−4	−1911	−18369	−10920

49—*continued*

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
15	+25	−1685	−19935	−24083
16	+56	−1384	−20524	−36336
17	+89	−1003	−19919	−46461
18	+124	−537	−17889	−53016
19	+161	+19	−14189	−54321
20	+200	+670	−8560	−48444
21	+241	+1421	−729	−33187
22	+284	+2277	+9591	−6072
23	+329	+3243	+22701	+35673
24	+376	+4324	+38916	+95128

9,800 167,230,700 74,451,107,640
1,566,040 12,408,517,940
1 1 $\frac{5}{6}$ $\frac{7}{12}$ $\frac{7}{60}$

(cont'd on next page)

TABLE 6-6 (Cont'd)

50

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
1	−104	−312	+96876	+10764
3	−103	−931	+93771	+31809
5	−101	−1535	+87631	+51419
7	−98	−2114	+78596	+68684
9	−94	−2658	+66876	+82764
11	−89	−3157	+52751	+92917
13	−83	−3601	+36571	+98527
15	−76	−3980	+18756	+99132
17	−68	−4284	−204	+94452
19	−59	−4503	−19749	+84417
21	−49	−4627	−39249	+69195
23	−38	−4646	−58004	+49220
25	−26	−4550	−75244	+25220
27	−13	−4329	−90129	−1755
29	+1	−3973	−101749	−30305
31	+16	−3472	−109124	−58652
33	+32	−2816	−111204	−84612
35	+49	−1995	−106869	−105567
37	+67	−999	−94929	−118437
39	+86	+182	−74124	−119652
41	+106	+1558	−43124	−105124
43	+127	+3139	−529	−70219
45	+149	+4935	+55131	−9729
47	+172	+6956	+125396	+82156
49	+196	+9212	+211876	+211876
41,650		770,715,400		372,255,538,200
	433,160		372,255,538,200	
2	$\frac{1}{2}$	$\frac{5}{8}$	$\frac{15}{128}$	$\frac{7}{30}$

51

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
0	−650	0	+21060	0
1	−647	−324	+20736	+7452
2	−638	−643	+19771	+14582
3	−623	−952	+18186	+21077
4	−602	−1246	+16016	+26642
5	−575	−1520	+13310	+31009
6	−542	−1769	+10131	+33946
7	−503	−1988	+6556	+35266
8	−458	−2172	+2676	+34836
9	−407	−2316	−1404	+32586
10	−350	−2415	−5565	+28518
11	−287	−2464	−9674	+22715
12	−218	−2458	−13584	+15350
13	−143	−2392	−17134	+6695
14	−62	−2261	−20149	−2870
15	+25	−2060	−22440	−12848
16	+118	−1784	−23804	−22616
17	+217	−1428	−24024	−31416
18	+322	−987	−22869	−38346
19	+433	−456	−20094	−42351
20	+550	+170	−15440	−42214
21	+673	+896	−8634	−36547
22	+802	+1727	+611	−23782
23	+937	+2668	+12596	−2162
24	+1078	+3724	+27636	+30268
25	+1225	+4900	+46060	+75670
11,050		221,375,700		47,861,426,340
	17,218,110		17,803,525,740	
1	3	$\frac{5}{6}$	$1\frac{7}{12}$	$\frac{3}{40}$

52

ξ'_1	ξ'_2	ξ'_3	ξ'_4	ξ'_5
1	−225	−135	+1620	+2700
3	−223	−403	+1572	+7988
5	−219	−665	+1477	+12943
7	−213	−917	+1337	+17353
9	−205	−1155	+1155	+21021
11	−195	−1375	+935	+23771
13	−183	−1573	+682	+25454
15	−169	−1745	+402	+25954
17	−153	−1887	+102	+25194
19	−135	−1995	−210	+23142
21	−115	−2065	−525	+19817
23	−93	−2093	−833	+15295
25	−69	−2075	−1123	+9715
27	−43	−2007	−1383	+3285
29	−15	−1885	−1600	−3712
31	+15	−1705	−1760	−10912
33	+47	−1463	−1848	−17864
35	+81	−1155	−1848	−24024
37	+117	−777	−1743	−28749
39	+155	−325	−1515	−31291
41	+195	+205	−1145	−30791
43	+237	+817	−613	−26273
45	+281	+1515	+102	−16638
47	+327	+2303	+1022	−658
49	+375	+3185	+2170	+23030
51	+425	+4165	+3570	+55930
46,852		162,342,180		26,358,466,680
	2,108,340		108,228,120	
2	1	$\frac{2}{3}$	$\frac{1}{24}$	$\frac{1}{20}$

Table 6-6 is taken from Table XXIII of Fisher and Yates: *Statistical Tables for Biological, Agricultural, and Medical Research*, published by Longman Group Ltd., London, (1974) 6th edition (previously published by Oliver & Boyd Ltd., Edinburgh) and by permission of the authors and publishers.

TABLE 6-7

EXAMPLE OF ORTHOGONAL POLYNOMIALS

x	P_1 $(-4+x)$	P_2 $(12-8x+x_2)$	P_3 $(-6+\frac{41}{6}x$ $-2x^2+\frac{1}{6}x^3)$	P_1P_2	P_1P_3	P_2P_3
1	−3	5	−1	−15	3	−5
2	−2	0	1	0	−2	0
3	−1	−3	1	3	−1	−3
4	0	−4	0	0	0	0
5	1	−3	−1	−3	−1	3
6	2	0	−1	0	−2	0
7	3	5	1	15	3	5
Total	0	0	0	0	0	0

TABLE 6-8

EXAMPLE 6-4 COMPUTATIONS

Barrel Length, in.	Sum of Velocities s_i, ft/s	Difference of Velocities d_i, ft/s	ξ_1'	ξ_2'	ξ_3'
18, 16	2183	13	1	−35	− 7
20, 14	2188	12	3	−29	−19
22, 12	2181	11	5	−17	−25
24, 10	2170	12	7	1	−21
26, 8	2142	8	9	25	− 3
28, 6	2124	44	11	55	33

Part of Table 6-8 is taken from Table XXIII of Fisher & Yates: *Statistical Tables for Biological, Agricultural, and Medical Research,* published by Longman Group Ltd., London, (1974) 6th edition (previously published by Oliver & Boyd Ltd., Edinburgh) and by permission of the authors and publishers.

TABLE 6-9

ANOVA TABLE FOR EXAMPLE 6-4

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F Ratio
Linear Regression	1	967.72	967.72	24.21
Quadratic Regression	1	1607.33	1607.33	40.20
Cubic Regression	1	65.80	65.80	1.65
Residual Error	8	319.82	39.98	
	11	2960.67		

Note that the SS are found from

$$(744)^2/572 = 967.72,\ (-4394)^2/12{,}012 = 1607.33, \text{ etc.}$$

Since quadratic regression is highly significant, but cubic regression is not, we fit the quadratic, which, in terms of the original values of the x, becomes

$$z = 1082.33 + 1.3007(2)(x - 17)/2 - 0.3658(3)\,[(x - 17)^2/4 - 143/12]$$

$$= 994 + 10.63x - 0.2744x^2, \text{ as in Example 6-3.}$$

The advantageous use of orthogonal polynomials in least squares curve fitting for numerous applied problems is clearly seen, especially along with significance tests for the coefficients in the form of an ANOVA as illustrated in Table 6-9.

For a generalized application of least squares principles woven into the problems of imprecision and inaccuracy of measurement discussed in Chapter 2, see the appendix to this chapter on the sampling of atmospheric ozone concentrations.

6-11 MULTIPLE REGRESSION OR THE GENERAL LINEAR MODEL

6-11.1 INTRODUCTION

Although we have discussed linear regression or linear least squares, the fitting of a plane or one dependent variable on two first order variables, the fitting of a dependent variable to three variables of the first power, the fitting of a quadratic, or a cubic, etc., we actually are performing the task of multiple linear regression. This is also recognized as fitting the "general linear" model. In the case of equal spaces on the abscissa, we were able to use orthogonal polynomials for swift fitting and were even able to develop stopping rules by using the ANOVA technique, i.e., appropriate statistical tests of significance. Thus it may be seen that least squares based on the general linear model represents a very powerful tool to employ in applications. There is, nevertheless, the problem of how many linear terms to use and where the general linear model should stop for an appropriately useful, simple, and compact equation, or "law", for any possible future use. A tremendous background of statistical material on the multiple linear regression problem exists, and the reader should consult any standard text on the subject, such as Mood and Graybill (Ref. 18).

We will give a very brief account of the general multiple linear regression problem, perhaps useful to the Army statistician, so that he may have a quick reference to accompany this chapter. Since we will be dealing with any number of independent variables or linear terms, it is urgent to resort to vector and matrix notation for these general solutions. Such an account clearly will fit well with general calculations on electronic computers too.

6-11.2 THE GENERAL LINEAR REGRESSION MODEL

We will consider as many as r independent (linear or other) variates x, which are free of error and for which there are n sample observations on each and corresponding measurements for the dependent variable y. Thus the independent variates x will be represented by the symbols

$$x_{ij},\ i = 1, \ldots, n \text{ and } j = 1, \ldots, r\ .$$

For the ith measurement of the jth independent variable, i.e., x_{ij}, there is a corresponding observed value of y, i.e., y_i, say $i = 1,2, \ldots, n$, which is subject to error.

Suppose we let

$$\beta_j \text{ for } j = 1,2, \ldots, r$$

represent the true unknown coefficients of the linear regression terms and define

$$e_i \text{ for } i = 1,2, \ldots, n$$

for the errors in the dependent variables y.

The basic linear regression model is then

$$y_i = \sum_{j=1}^{r} \beta_j x_{ij} + e_i^*,\ \ i = 1,2, \ldots, n \tag{6-133}$$

for which we will fit the linear relation by the method of least squares

$$y = \sum_{j=1}^{r} b_j x_j^* \tag{6-134}$$

where the b_j are the "best" estimates of the β_j.

---

*The reader should note that we are using a rather general form, and to illustrate an intercept or have a constant term, say β_0, for example, we could employ the sum indicated by $\sum_{i=0}^{r-1} \beta_j x_{ij} + e_i$ as the model.

As is usual and for use in significance tests, we will assume that the errors e_i are normally distributed with mean value zero and common variance σ^2, i.e.,

$$e_i \rightarrow N(0,\sigma^2). \tag{6-135}$$

Also we have that

$$E(y_i) = \sum_{j=1}^{r} \beta_j x_{ij} \tag{6-136}$$

and

$$\mathrm{Var}(y_i) = \sigma^2.$$

It will be convenient, in view of the need for a constant term, to designate often that

$$x_{i1} = 1 \quad \text{for all } i.$$

We use the brackets [ ] to designate a vector or matrix, as the case may be; then we may define

$$[y] = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \tag{6-137}$$

$$[x] = \begin{bmatrix} x_{11}\ x_{12} \ldots x_{1r} \\ x_{21}\ x_{22} \ldots x_{2r} \\ \cdot \qquad\qquad \cdot \\ \cdot \qquad\qquad \cdot \\ \cdot \qquad\qquad \cdot \\ x_{n1}\ x_{n2} \ldots x_{nr} \end{bmatrix} \tag{6-138}$$

$$[\beta] = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_r \end{bmatrix} \tag{6-139}$$

$$[b] = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_r \end{bmatrix} \tag{6-140}$$

$$[e] = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{bmatrix} \tag{6-141}$$

With these vector and matrix notations, the system of linear equations (Eq. 6-133) becomes simply

$$[y] = [x]\,[\beta] + [e] \tag{6-142}$$

which in effect needs solution for the β_i. The β_i are estimated by the b_i, which are determined by the method of least squares.

It is well-known (see, for example, Ref. 18) that the least squares estimates $b_i = \hat{\beta}_i$ of the β_i are determined from

$$[b] = [\hat{\beta}] = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\beta}_r \end{bmatrix} = \{[x]^T[x]\}^{-1}[x]^T[y] \tag{6-143}$$

or the vector solution of the $\hat{\beta}$'s is found by inverting the product of the transpose of the matrix of the x's with the matrix of the x's, and this product is multiplied by the transpose of the matrix of the independent variables and the vector $[y]$ of the dependent observations.

Since there are many computer programs on file to multiply and invert nonsingular matrices, the solution of Eq. 6-143 for any number of unknowns is readily adapted to high-speed computation.

It can be shown that the b_i or $\hat{\beta}_i$ are consistent, efficient, sufficient, and minimum variance unbiased estimates of the true β_i for the model (Eq. 6-133).

It can also be shown that the residual variance σ^2 is estimated from

$$\hat{\sigma^2} = \{[y] - [x]\,[\hat{\beta}]\}^T\{[y] - [x]\,[\hat{\beta}]\}/(n-r) \tag{6-144}$$

which also is easily calculated on a computer.

The quantity

$$(n-r)\hat{\sigma^2}/\sigma^2 = \chi^2(n-r) \tag{6-145}$$

follows the chi-square distribution with $(n-r)$ df.

The covariance matrix of the estimators of the β's is given simply by the quantity

$$\mathrm{Cov}[\hat{\beta}] = \sigma^2\{[x]^T[x]\}^{-1}. \tag{6-146}$$

Finally, the estimators $\hat{\beta}_i$ of the true coefficients and the estimator $\hat{\sigma^2}$ of the variance of residuals are distributed independently in probability. Moreover, the vector $[\hat{\beta}]$ follows an r-variate normal distribution with mean equal to $[\beta]$ and covariance given by Eq. 6-146.

With regard to confidence intervals on the unknown coefficients β_i, suppose we let c_{ij} represent the ijth element of the inverse matrix $[C]$ defined as

$$[C] = \{[x]^T[x]\}^{-1}. \tag{6-147}$$

Then $(1-2\gamma)$ confidence bounds on the β_i's individually—but not all jointly—may be determined from the probability statement

$$Pr\left[\hat{\beta}_i - t_\gamma\sqrt{c_{ii}\hat{\sigma^2}} < \beta_i < \hat{\beta}_i + t_\gamma\sqrt{c_{ii}\hat{\sigma^2}}\right] = 1 - 2\gamma. \tag{6-148}$$

The confidence bounds of Eq. 6-148 on any of the β_j in Eq. 6-136 are of considerable use in applied least squares or regression analyses, although the physical scientist and statistician will surely have more direct interest in overall confidence statements about the entire hyperplane or linear model (Eq. 6-136). Fortunately, such a confidence type of statement is possible because of some pioneering results of Henry Scheffé (Ref. 3). In fact they represent an extension of Scheffé's results for the fitted line, as in Eq. 6-26, which uses the Fisher-Snedecor F statistic.

Recently, Taylor and Moore (Ref. 19) have added to inferences from Scheffé's results (Ref. 3) for the general linear model, which is our prime interest.

We will record some of the key results, which should be of value in many Army applications, and these apply mainly to the (whole) fitted line (Eq. 6-134) or a polynomial of degree $(r-1)$. We will illustrate these two cases comparatively by the following definitions:

Case I—Let the row vector $[X]_0^T$ be the linear form

$$[X]_0^T = [1, x_1, x_2, \ldots, x_{r-1}] \tag{6-149}$$

where we have taken $x_0 = 1$, say.

Alternatively, let us consider also the possibility of Case II—Let the row vector $[X]_0^T$ be the polynomial form

$$[X]_0^T = [1, x, x^2, \ldots, x^{r-1}]. \tag{6-150}$$

Note that we are now using a capital X to represent either the linear form in independent variables, such as in Eq. 6-134, the polynomial form, such as for the row vector of Eq. 6-150, or we could use it to represent any sum of mathematical terms.

For the row vector of linear terms in Eq. 6-149, the observed values of the independent variables take the form of the matrix $[X]$ in Eq. 6-138. On the other hand, for the polynomial fit of Eq. 6-150, the observed values of the independent variables may be represented schematically as the matrix

$$[X] = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{r-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{r-1} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ 1 & x_n & x_n^2 & \ldots & x_n^{r-1} \end{bmatrix} \tag{6-151}$$

In terms of the linear form of independent variables in Eq. 6-149, Eqs. 6-133 and 6-134 still hold, of course. Also the least squares estimates of the β_i, or the $\hat{\beta}_i$, are given by Eq. 6-143. These statements merely represent a review for the purpose of leading up to and recording that the fitting of a polynomial—or actually any other sum of terms—is not different from fitting the ordinary linear terms. In fact, it is seen that the estimates of the β_i for the polynomial are given by the matrix manipulations

$$[\hat{\beta}] = \{[X]^T[X]\}^{-1}[X]^T[y] \tag{6-152}$$

which is the same form or result as in Eq. 6-143 for linear terms.

Note also that

$$[y] = [X]\,[\beta] + [e] \tag{6-153}$$

represents all of the n equations, given for a general $i = 1,2, \ldots, n$ by

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_{r-1} x_i^{r-1} + e_i \tag{6-154}$$

or the true polynomial plus an error e_i.

Continuing, the variance of residuals is still estimated by Eq. 6-144 or in this case for the polynomial (Eq. 6-154) by

$$\hat{\sigma}^2 = \{[y] - [X][\hat{\beta}]\}^T\{[y] - [X][\hat{\beta}]\}/(n-r) \tag{6-155}$$

and confidence bounds on (or significance tests of) any individual, unknown coefficient β_i may be determined with the aid of the schematic form of Eq. 6-148.

We can now make confidence statements about the entire fitted line, or a polynomial, or even a general linear form by using the unique theorem of Scheffé (Ref. 3). As an example, consider any selected value of x, say x_0, representing a point of interest in the line or fitted curve. Then for example, for the polynomial, and the row vector

$$[X_0]^T = (1, x_0, x_0^2, \ldots, x_0^{r-1}), \tag{6-156}$$

Taylor and Moore (Ref. 19) show that the quantity

$$\hat{\sigma}^2[X_0]^T\{[X]^T[X]\}^{-1}[X_0] \tag{6-157}$$

gives an unbiased estimate of the variance of prediction from the fitted curve, and moreover, $(1 - 2\gamma)$ confidence bounds on the value y_0 predicted from x_0 are determined from

$$\begin{aligned}[X]^T[\hat{\beta}] - [rF_\gamma(r, n-r)]^{1/2}\,\hat{\sigma}\sqrt{[X_0]^T\{[X]^T[X]\}^{-1}[X_0]} \\ \leq [X_0]^T[\hat{\beta}] \leq \\ [X]^T[\hat{\beta}] + [rF_\gamma(r, n-r)]^{1/2}\,\hat{\sigma}\sqrt{[X_0]^T\{[X]^T[X]\}^{-1}[X]}\end{aligned} \tag{6-158}$$

where

F_γ = upper γ probability level of F with r and $(n-r)$ df.

For the line the reader may check that Eq. 6-154 reduces to

$$\hat{\beta} + \hat{\beta}_1 x_0 \pm [2F_\gamma(2, n-2)]^{1/2}\,\hat{\sigma}[(1/n) + n(x_0 - \bar{x})^2/A_{xx}]^{1/2} \tag{6-159}$$

which is equivalent to the result given in par. 6-2.2.

In addition to the determination of confidence bounds or regions for a polynomial fit to the data, Taylor and Moore (Ref. 19) also discuss the two-sample and the k-sample cases for the linear and polynomial fits to the original data, along with the appropriate pooled variance of residuals and establishment of confidence bounds on the curve fitted. Thus this should represent some likely applications the Army analyst could use.

Christensen (Ref. 20) also discusses simultaneous statistical inference for the normal multiple linear regression model from the standpoint of Scheffe's use of the F-tests and the Bonferroni t-tests, but neither is uniformly superior. If regressors can be controlled to be uncorrelated, the Bonferroni t-tests are superior.

Breaux (Ref. 21) covers the subjects of "stepwise" multiple linear regression and the use of computers to fit curves at various stopping points. Initially, one may have only a hazy idea about the actual type of "law" he will fit to the data; therefore, stepwise procedures could be of considerable value.

Clearly, this account of multiple linear regression or the general linear model brings up the very important question of which useful "law" should be fitted to the data taken and the critical point of where to stop if we are trying to use a number of terms that will best represent our final judgment, perhaps for future prediction purposes. With regard to which law to fit, it seems that this should depend on the physics or engineering of the situation; otherwise the statistician may be criticized for a "blind" fit. On the other hand, the demands of time

could call for a quick fit, which the statistician could develop as a "stopgap" rule or law, especially since the physicist or engineer might take too long, even years, to develop a perfectly acceptable law. Thus good insight and judgment are often called for in Army applications of curve fitting.

Stopping rules very often apply to the "statistical" type of fit, and a number of papers on the subject have been published. A natural approach is to use a standard Student's t test for each coefficient of a term that is added to determine whether that particular coefficient differs significantly from zero. If the coefficient is not significantly different from zero, the corresponding term is not included, whereas if it does indeed differ significantly from zero, then the term is included. Effroymson (Ref. 22) recommends the use of Student's t or F type of test involving correlation and partial correlation coefficients of the next fitted term and gives very specific rules for the inclusion or rejection of that particular term. Also Forsythe, Engelman, Jennrich, and May (Ref. 23) recommend the use of a permutation type of test that offers a stopping rule for "forward stepping". Perhaps these references will be of some value to the analyst who is required to make such Army applications toward obtaining a useful fit to the data.

6-12 FUNCTIONAL RELATIONS AND NONLINEAR REGRESSION OR GENERALIZED LEAST SQUARES (WITH OR WITHOUT ERROR IN INDEPENDENT VARIABLES)

6-12.1 INTRODUCTION

So far, we have covered primarily the problem of "linear" least squares or regression and with some account of its relation to the use of physical laws in practice. Also we have shown how "linear" regression extends easily to nonlinear forms. Our purpose has been to indicate a rather compact approach through the use of the A_{uv}-type computations or functions in the analysis and to show that in practice it is usually, or in many cases, highly desirable to work with physical relations or parameters, if at all possible, since such models are more informative, physically meaningful, and will be more enduring and of wider interest. It is, nevertheless, clear that we cannot begin to cover such an involved and wide field of interest in any depth here. In fact, the important objective of finding the most appropriate use or combination of statistical methods with models or laws in the physical sciences represents a field of interest that is always undergoing development. The best gains will likely result in bridging the gap between the science of statistics on one hand and the field of physical application on the other. Nonlinear or generalized least squares, with or without errors in the independent variables, is therefore a wide-open field that critically depends on particular applications. However, the decision to fit a hypothesized or developed model for the particular problem at hand seems to lie most frequently outside the normal judgment of the analyst or practicing statistician and often is dictated by the physical application or by a nonstatistician with much expertise otherwise, who works full time in a given field of application. Hence the need for a team effort and continual cross-fertilization of statistical principles with the physical sciences to develop superior, or even most useful, results. Thus we will have to limit our account to an introduction, a few principles, and some pertinent references to some of the current literature on the general subject.

Initially, we will frequently encounter a variety of applications for which there will be observational errors in both the independent variable(s) and the dependent variable, so that the right-hand sides (RHS) of Eqs. 6-49 and 6-50 will apply, especially for the situation when η is not a linear function of μ. More concretely, the basic model might be represented as

$$x = \mu + e \tag{6-160}$$

$$y = \eta + d = f(\mu) + d. \tag{6-161}$$

Hence we can say that our primary problem is either to determine the best relation between μ and η, i.e., to determine

$$\eta = f(\mu) \tag{6-162}$$

or to hypothesize from physical considerations some appropriate relation (Eq. 6-162), and then to judge statistically whether the fitted law is suitable for general use. This means that we will be able, through calculations or appropriate iterations, to weed out the effects of the errors d and e.

The physical law represented by Eq. 6-162 may relate, for example, to the penetration of armor, flight characteristics of a new projectile in terms of its key parameters, a stress-strain diagram, or even the validity of Lanchester's square law for the estimation of battle casualties. In our example of Fig. 6-1 and the data of Table 6-2, we selected fitting the residual energy on the striking energy of the projectiles as perhaps the "best" law, which also gave a rather simple method of calculating confidence bounds on the critical velocity. We found, in fact, that rather tight confidence bounds could be found from this procedure. Of course, we must clearly explain that applications are such that often not even a single law will exist that is applicable, and the investigator may have to be very clever indeed to find the most appropriate, or even a very useful, relationship between parameters of major interest when he is fitting curves to observational data. Also it is fortunate and often true that any one of several selected laws might be sufficient in any single application, at least as a "stopgap" procedure at the time and until the more appropriate physical rule can be found.

Although it cannot always be guaranteed, it is, nevertheless, a very good and useful rule to control the independent variables at important or key levels of interest so that they can be relatively free of error insofar as the regression analysis is concerned. Naturally, if all the independent variables are relatively free of error compared to the dependent variable of interest, the curve fitting problem would be simplified. However, if all of the independent variables do contain errors—the relative sizes of which are unknown—we face the more general and difficult problem. We believe that the best choice of topics here would be to indicate two useful algorithms for the nonlinear or generalized least squares problem—one is the case in which the independent variables are entirely free of error, studied by Gallant (Ref. 24); and the other is the outline of the complex case covering errors in all of the variables, both dependent and independent, studied by Britt and Luecke (Ref. 25) and others.

6-12.2 THE GALLANT ALGORITHM (ERROR-FREE INDEPENDENT VARIABLES)

For the case of nonlinear regression with error in the dependent variable only and a number of independent variables and parameters of interest, Gallant (Ref. 24) considers a generalization of Eqs. 6-160 and 6-161 with the letters now representing vectors or matrices but with the errors in the x's, i.e., e, all equal to zero. In other words, he considers the case

$$[y] = [f(x,\mu)] + [d]^*, \qquad (1 \times n) \tag{6-163}$$

where the quantity $[y]$ is a vector of dependent variables with

$$[y]^T = [y_1, y_2, \ldots, y_n], \qquad (n \times 1) \tag{6-164}$$

the n observations on the dependent variable subject to errors

$$[d]^T = [d_1, d_2, \ldots, d_n], \qquad (n \times 1) \tag{6-165}$$

and the function f, the best known physical relation between y and the independent variables x ($=\mu$ in this case), which is represented by the vector function of observations

$$[f(x,\mu)] = [f(x_1,\mu), f(x_2,\mu), \ldots, f(x_n,\mu)], \qquad (n \times 1) \tag{6-166}$$

and $[\mu]$ is the unknown vector of p parameters

*Actually, for the case in which $e_i = 0$, the first vector on the RHS may be written as $f(\mu)$ since the independent variable x attains its true value μ.

$$[\mu] = [\mu_1, \mu_2, \ldots, \mu_p], \qquad (p \times 1) \tag{6-167}$$

to be estimated for the functional form fitted.

The SS of deviations of the observed values of y minus the fitted function f corresponding to estimated values of the parameters $[\mu]$ is given by

$$\mathrm{SSE}(\mu) = \Sigma(y - f)^2 = [y - f(\mu)]^T[y - f(\mu)] \tag{6-168}$$

Eq. 6-168 being in vector form.

By analogy with the linear form of Eqs. 6-49 and 6-50, we might say in the generalized nonlinear regression problem that the function f replaces the linear term of Eq. 6-50, otherwise serving the same purpose, but that to find the appropriate or best fit of the function f, we have to carry out an iteration process. Ordinarily, this type of iteration is done by the so-called Gauss-Newton method, or Hartley's modified Gauss-Newton technique (Ref. 26), or by Marquardt's algorithm (Ref. 27). The Gauss-Newton method usually is based on the substitution or first-order approximation of a Taylor series expansion of the fitted or response function f in the equation for the SS for error. This means that the Taylor series expansion is truncated at the term involving first derivatives of the function f with respect to the p unknown parameters $[\mu]$. Thus we designate the $n \times p$ matrix of derivatives with respect to the parameters given by Eq. 6-167 as

$$[f'(\mu)] = \left[\frac{\partial}{\partial \mu_j} f(x_i, \mu)\right]^* \tag{6-169}$$

where i indicates row index, and j indicates the column index—and calculate this matrix of derivatives for use in the iteration process.

As pointed out by Gallant (Ref. 24), the iteration that determines the final fit, or the algorithm, proceeds as follows:

(0) Choose a starting estimate $[\mu_0]$ of the unknown vector $[\mu]$, and compute

$$[D_0] = \{[f'(\mu_0)]^T[f(\mu_0)]\}^{-1}[f'(\mu_0)]^T[y - f(\mu_0)]. \tag{6-170}$$

Then find a λ_0 between 0 and 1 such that

$$\mathrm{SSE}(\mu_0 + \lambda_0 D_0) \leq \mathrm{SSE}(\mu_0). \tag{6-171}$$

(1) Let $\mu_1 = \mu_0 + \lambda_0 D_0$. Next compute

$$[D_1] = \{[f'(\mu_1)]^T[f(\mu_1)]\}^{-1}[f'(\mu_1)]^T[y - f(\mu_1)]. \tag{6-172}$$

Then find a λ_1 between 0 and 1 such that

$$\mathrm{SSE}(\mu_1 + \lambda_1 D_1) \leq \mathrm{SSE}(\mu_1). \tag{6-173}$$

(2) Let $\mu_2 = \mu_1 + \lambda_1 D_1$, and then proceed with the same type of calculation as before, i.e., as in Eq. 6-172, except that μ_1 is replaced by μ_2. This iterative process is repeated through the number of steps required to make the difference between μ_i at the ith stage and μ_{i+1} at the $(i+1)$st stage as small as desired, for example, to some number of decimal places, and also to make the difference between the SS of error at the ith and $(i+1)$st stages suitably small. Hartley (Ref. 26) gives two very useful methods for choosing the step length λ_i.

If the size of the sum of squares of errors is too large, one should consider that the chosen function f is perhaps not the best one for this particular problem. Hence consideration should be given to another choice.

*This matrix is a Jacobian of the quantities.

As might be expected, the improved choice may depend upon extensive familiarity with the field of application.

Finally, estimates of the parameters in the vector $[\hat{\mu}]$ converge almost surely to the true unknown vector $[\mu]$, and the quantity given by

$$\sqrt{n}\{[\hat{\mu}] - [\mu]\} \tag{6-174}$$

converges in distribution to a p-variate normal type of frequency function with mean $[\mu]$ and the variance-covariance matrix given by

$$\sigma^2\{(1/n)\,[f'(\hat{\mu})]^T[f'(\hat{\mu})]\}^{-1}. \tag{6-175}$$

Obviously, the fitting of nonlinear least squares to experimental data can become complex indeed, and due to the nature of the rather extensive computations, it seems best to program the calculations on appropriate electronic computers.

Again, we remark that proper choice of the best function to fit continues to deserve special attention.

6-12.3 THE BRITT AND LUECKE ALGORITHM FOR ESTIMATING PARAMETERS IN NONLINEAR MODELS WITH ERRORS IN BOTH THE DEPENDENT AND INDEPENDENT VARIABLES*

For the case of fitting a general functional relationship to observed data when both the dependent and independent variables are subject to errors and several parameters in the nonlinear function must be estimated, the fitting process by least squares becomes very involved. Again, an iterative computational procedure is necessary to make the adjustment. Historically, this has been one of the more important topics in the physical sciences and mathematical statistics. In 1943 Deming (Ref. 28) published a book titled the *Statistical Adjustment of Data*, which is devoted primarily to this subject. The algorithm developed by Deming (Ref. 28) may still be of interest as a useful input to the procedure of Britt and Luecke, which we outline here.

For the much simpler case of no errors in the independent variables, one must experience the application of only a linear form in the parameters to obtain direct solutions to the least squares problem. When the functional relationship to be fitted is nonlinear or even for fitting a line with errors in both dependent and independent variables, iterative procedures are needed except in the most special cases. Fortunately, Britt and Luecke's (Ref. 25) development is general enough to include practically all such problems. Therefore, we will outline their procedure since it is perhaps the more useful and important one for most Army applications.

The algorithm of Britt and Luecke (Ref. 25) covers a much different approach compared to that we have discussed so far; it does not split up the dependent and independent variables into separate vectors. Rather, they use a vector z, which includes all of the "observables", i.e., including all observations on both the dependent and independent variables. (The reader should note here that we simply have used the letter z for a vector. The use of brackets for all vectors or matrices in this particular numbered subparagraph would be very cumbersome. Hence all letters, Arabic and Greek, and functions alike will denote vectors or matrices in our presentation of this numbered subparagraph.) Thus the vector z is a "long" vector and includes all of the observed values of both the dependent and independent variables considered in the least squares adjustment procedure. It is a matter, therefore, of keeping the components of the vector "straight". The functional form fitted, or the vector function designated as $f(z,\theta)$, should be "well-behaved" in the region of interest. The Britt and Luecke algorithm (Ref. 25) develops a technique that gives ML estimates of the unknown parameters for the assumption covering normally distributed errors of measurement along with "known" variances and covariances for the errors. Most often, the error variance-covariance matrix will not be known, and some appropriate estimate of it will have to be assumed. In this case the resulting estimates of the parameters cover what is called a "weighted least squares" adjustment. This causes no essential change in the problem. Insofar

*To avoid unusually cumbersome notation in this paragraph, we have not used brackets for vectors. Arabic and Greek letters, and function symbols are to be understood as representing vectors or matrices.

as is possible we will use the notation of Britt and Luecke (Ref. 25) although we have already used n for the number of observations on the dependent and independent variables. Accordingly, another letter must be used in this discussion. Their algorithm considers a p vector (they use "n" instead of "p") of unknown parameters θ_0 to be estimated for the functional relationship; a q vector of all of the dependent and independent variable observations, or observables; and a k vector of functional forms $f(z,\theta)$, which are used with the form or property

$$f(z_t,\theta_0) = 0. \tag{6-176}$$

The subscript "t" on z is used to designate the true value of the observables, whereas "0" is used as a subscript for the θ to distinguish it from a step "i" in the iteration, i.e., θ_i. Thus the measurements of the true z_t contain random experimental errors; therefore, the measurements are represented as

$$z_m = z_t + e \tag{6-177}$$

where z_m is the q vector of measurements, and the quantity e is a q vector of the errors of the dependent and all independent variables. The reader should note that actually the vector z_t is also a vector of unknown true values of the dependent and independent variables, which are to be estimated also. Hence during the entire iteration process, both the parameters and the true values of the z's will be estimated in the Britt-Luecke algorithm. For the iterative process there are a number of conditions that must be satisfied, as pointed out by Britt and Luecke (Ref. 25). They are

1. The function f is continuous.
2. The partial derivatives of $f=f(z,\theta)$ with respect to both arguments, z and θ, exist and are continuous.
3. The second partial derivatives of each component of the vector function f with respect to both arguments exist and are bounded.
4. The $k \times p$ Jacobian matrix, call it $f_\theta = f_\theta(z,\theta)$, of partial derivatives of f with respect to θ has rank p. The $k \times q$ Jacobian matrix, call it $f_z = f_z(z,\theta)$, of the function f has rank k.

The vector of errors e for the dependent and independent variables is assumed to follow a multivariate normal distribution with mean values equal to zero and to possess a known positive definite variance-covariance matrix R, i.e.,

$$E(e) = 0 \tag{6-178}$$

and

$$E(ee^T) = R.$$

The algorithm of Britt and Luecke (Ref. 25) involves, as before, a truncated Taylor series expansion of the function f and the use of a k vector of Lagrange multipliers to obtain the minimization required. We will summarize the final results for any possible Army applications, and otherwise interested readers may consult the Britt-Luecke paper (Ref. 25).

First, the vector giving the difference between the true parameters and the values at the ith iterative stage is a p vector represented by the following for a selected or fitted function f:

$$\theta - \theta_i = -[f_\theta^T(f_z R f_z^T)^{-1} f_\theta]^{-1} f_\theta^T (f_z R f_z^T)^{-1}[f(z_i,\theta_i) + f_z(z_m - z_i)] \ . \tag{6-179}$$

As usual, one starts with the measured values or observables z_m and initial estimates of the parameters for the first stage $i = 1$, with also in this algorithm initial estimates of the true values z_t for the first iteration. Both the estimates of the parameters and the true z values may be taken from known experience, the physical situation (if that knowledge is available), from a preliminary study of the problem, or even perhaps determined from a least squares fit of a "linearized" form if one can be obtained. Similar considerations will apply to the variance-covariance matrix of errors, or one may use different inputs to judge the sensitivity of the variance-covariance matrix to the estimation of parameters.

The true values of the dependent and the independent variables are determined from iterations indicated in Ref. 25

$$z - z_i = z_m - z_i - Rf_z^T(f_z R f_z^T)^{-1}[f(z_i,\theta_i) + f_\theta(\theta - \theta_i) + f_z(z_m - z_i)]. \quad (6\text{-}180)$$

Although the vector of Lagrange multipliers is of no direct interest here, interested readers may calculate this vector by an iterative process given in Eq. 24 of the Britt-Luecke paper (Ref. 25).

As is customary in standard iterative problems of the kind discussed here, one stops at that particular step for which his calculated value at the stage differs only by an appropriate smallness criterion with the preceding step of iteration. Again, we mention that another reasonable physical model might be used if necessary to obtain the best adjustment for prediction purposes.

The variance-covariance matrix for the estimation errors of the parameters is given by Britt and Luecke in Ref. 25 as

$$E[(\hat{\theta} - \theta_0)\,(\hat{\theta} - \theta_0^T)^T] = [f_\theta^T(f_z R f_z^T)^{-1} f_\theta]^{-1} \quad . \quad (6\text{-}181)$$

An example of the fitting process is given by Britt and Luecke in Ref. 25, which uses data formerly analyzed by Deming (Ref. 28).

A number of other investigators have developed useful algorithms for the generalized least squares procedures with errors in dependent and independent variables—for example, the works of Dolby (Ref. 29), Celmins (Ref. 30), and Pope (Ref. 31). Celmins (Ref. 32) comments on the use of nonlinear least squares in the field of meteorological data experiments.

6-13 SUMMARY

In this chapter, we have covered a fairly wide range of topics in least squares, regression, and curve fitting in general. We have developed in detail the proposition that one should seek out not only the fitting of lines and polynomials to observational data, but, if at all possible, he should try to adjust physically meaningful models to the data at hand. In this way more enduring regression models may be recorded for prediction purposes.

We have covered both of the important cases in practice in which the independent variables sometimes may be free of errors of determination and the case most often met for which the independent variables are subject to error, as the dependent variable always is. Methods for the estimation of parameters for both cases have been covered, and several illustrative examples have been presented and discussed fully.

REFERENCES

1. AMCP 706-110, Engineering Design Handbook, *Experimental Statistics, Section 1, Basic Concepts and Analysis of Measurement Data,* December 1969.
2. *ASTM Manual on Fitting Straight Lines (1962)*, (ASTM Special Technical Publication No. 313), Published by the American Society for Testing and Materials, 1916 Race St., Philadelphia, PA 19103, 1962.
3. Henry Scheffe', *The Analysis of Variance*, Section 3.3, John Wiley and Sons, Inc., New York, NY, 1961.
4. Joseph Berkson, "Are There Two Regressions?", Journal of the American Statistical Association **45**, 164-80 (1950).
5. Abraham Wald, "Fitting Straight Lines If Both Variables Are Subject to Error", The Annals of Mathematical Statistics **11**, 284-300 (1940).
6. M. S. Bartlett, "Fitting A Straight Line When Both Variables Are Subject To Error", Biometrics **5**, 207-12 (1949).
7. Albert Madansky, "The Fitting of Straight Lines When Both Variables Are Subject to Error", Journal of the American Statistical Association **54**, 173-205 (1959).
8. Jerzy Neyman, "Existence of Consistent Estimate of the Directional Parameter in a Linear Structural Relation Between Two Variables", Annals of Mathematical Statistics, **22**, 496-512. (1951).

9. P. A. P. Moran, "A Test of Significance for an Identifiable Relation", Journal of the Royal Statistical Society, Series B, **18**, 62-4 (1956).
10. Frank E. Grubbs, "On Estimating Precision of Measuring Instruments and Product Variability", Journal of the American Statistical Association **43**, 243-64 (1948).
11. A. S. Goldberger, *Econometric Theory*, John Wiley and Sons, Inc., New York, NY, 1964.
12. Farrell E. Bloch, "Measurement Error and Statistical Significance of an Independent Variable", The American Statistician **32**, 26-7 (February 1978).
13. Allen H. Reed and Gaw T. Wu, "Estimation of Bias in Classical Linear Regression Slope When the Proper Model Is Functional Linear Regression", Communications in Statistics—Theory and Methods **A6(5)**, 405-16 (1977).
14. D. H. Richardson and D. Wu, "Least Squares and Grouping Method Estimators in the Errors in Variables Model", Journal of the American Statistical Association **65**, 724-48 (1970).
15. Edi Karni and Ishay Weissman, "A Consistent Estimator of the Slope in a Regression Model With Errors in the Variables", Journal of the American Statistical Association **69**, 211-3 (March 1974).
16. J. Neyman and E. L. Scott, "On Certain Methods of Estimating the Linear Structural Relation", Annals of Mathematical Statistics **22**, 352-61 (1951).
17. R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research*, Second Edition, Oliver and Boyd, Ltd., London, England, 1943.
18. A. M. Mood and F. A. Graybill, *Introduction to the Theory of Statistics*, Second Edition, McGraw-Hill Book Company, New York, NY, 1963.
19. M. S. Taylor and J. R. Moore, *Confidence Bounds for the General Linear Model*, Ballistic Research Laboratory Report No. 1986, May 1977.
20. Laurits R. Christensen, "Simultaneous Statistical Inference in the Normal Multiple Linear Regression Model", Journal of the American Statistical Association **68**, 457-61 (June 1973).
21. Harold J. Breaux, *On Stepwise Multiple Regression*, Ballistic Research Laboratory Report No. 1369, August 1967.
22. M. A. Effroymson, "Multiple Regression Analysis", A. Ralston and H. S. Wilf, Eds., *Mathematical Methods for Digital Computers*, Part V, (17), John Wiley and Sons, New York, NY, 1960.
23. A. B. Forsythe, L. Engelman, R. Jennrich, and P. R. A. May, "A Stopping Rule for Variable Selection in Multiple Regression", Journal of the American Statistical Association **68**, 75-7 (March 1973).
24. A. R. Gallant, "Nonlinear Regression", The American Statistician **29**, 73-81 (May 1975).
25. H. I. Britt and R. H. Luecke, "The Estimation of Parameters in Nonlinear, Implicit Models", Technometrics **15**, 233-47 (May 1973).
26. H. O. Hartley, "The Modified Gauss-Newton Method for the Fitting of Nonlinear Regression Functions by Least Squares", Technometrics **3**, 269-80 (1961).
27. D. W. Marquardt, "An Algorithm for Least Squares Estimation of Nonlinear Parameters", Journal of the Society For Industrial and Applied Mathematics **11**, 431-41 (1963).
28. W. E. Deming, *Statistical Adjustment of Data*, John Wiley and Sons, New York, NY, 1943.
29. G. R. Dolby, "Generalized Least Squares and Maximum Likelihood Estimation of Nonlinear Functional Relationships", Journal of the Royal Statistical Society, Series B, **34**, 393-400 (1972).
30. Aivars Celmins, *Least Squares Adjustment With Finite Residuals For Nonlinear Constraints and Partially Correlated Data*, Ballistic Research Laboratory Report No. 1658, July 1973.
31. Allen J. Pope, "Two Approaches to Nonlinear Least Squares Adjustments", The Canadian Surveyor **28**, 663-8 (December 1974).
32. Aivars Celmins, "Comments on 'Use of Nonlinear Least Squares in Meteorological Applications' ", Journal of Applied Meteorology **17**, 1579-80 (October 1978).

BIBLIOGRAPHY

R. L. Anderson and E. E. Houseman, *Tables of Orthogonal Polynomial Values Extended to N = 10"*, Research Bulletin 297, Iowa State University Agricultural Experiment Station, Ames, IA, 1942.

R. H. Bacon, "The 'Best' Straight Line Among the Points", American Journal of Physics **21**, 428-46 (1953).

George E. P. Box, *Fitting Empirical Data*, Technical Report No. 151, Army Mathematics Research Center, Madison, WI, May 1960.

H. A. David and Beverly E. Arens, "Optimal Spacing in Regression Analysis", The Annals of Mathematical Statistics **30**, 1072-81 (December 1959).

G. J. Hahn, "Fitting Regression Models With No Intercept Term", Journal of Quality Technology **9** (April 1977).

William Sachs, "Implicit Multifunctional Nonlinear Regression Analysis", Technometrics **18**, 161-73 (May 1976).

Charles P. Winsor, "Which Regression?", Biometrics **2**, 101-14 (December 1946).

APPENDIX 6A

A LEAST SQUARES APPLICATION TO PRECISION AND ACCURACY OF MEASUREMENT

6A-0 LIST OF SYMBOLS

A = constant

A' = constant term = $\bar{O}$ in Eq. 6A-5

a = constant

B, C, D = coefficients of linear, square, and cubic terms, respectively, when the altitudes are expressed in terms of orthogonal polynomials for least squares fits

B', C', D' = coefficients of linear, square, and cubic terms, respectively, of orthogonal polynomials in ξ'

b, c, d = coefficients of linear, square, and cubic terms, respectively, of a polynomial

d_i = difference of readings at altitude h_i

e_{ij} = random error of measurement of instrument $j = 1, 2, 3$ at altitude h_i

$\bar{h}$ = average altitude at which ozone measurements were made

h_i = ith altitude, km

$N(0,\sigma_{e_j})$ = denotes that the errors of measurement are normally distributed with zero mean and standard deviation σ_{e_j}

n = sample size

O_{ij} = observed ozone concentration at altitude h_i as measured by instrument j

$\delta O_{i-j} = \hat{\beta}_i - \hat{\beta}_j$ = estimate of difference in biases for the ith and jth instruments. The biases are considered to vary with altitudes.

s_i = sum of readings at altitude h_i

t = Student's t variate

α_j = denotes slope of a line if total instrumental bias can be modeled linearly

β_j = constant bias or systematic error of instrument j over the altitudes of interest. In one of the models the β_j are assumed to vary with altitude h_i. $j = 1, 2, 3$.

λ = coefficients chosen to give the orthogonal polynomial values ξ' whole numbers

ξ_i = ith order or power of the orthogonal polynomial

$\xi_i' = \lambda\xi_i$ = transformed orthogonal polynomial

σ_{av} = average imprecision of measurement for several similar instruments

$\hat{\sigma}_e$ = refers to a general standard error of measurement

$\hat{\sigma}_{e_1}, \hat{\sigma}_{e_2}, \hat{\sigma}_{e_3}$ = estimated standard deviations of errors of measurement for 1st, 2nd, and 3rd instruments, respectively

$\hat{\sigma}_{e_i - e_j}$ = estimated standard deviation of the difference in random errors of measurement of the ith and jth instruments

$\omega_i = \omega(h_i) = f(h_i)$ = true unknown ozone concentration at altitude h_i

6A-1 PRELIMINARY REMARKS

The use of least squares and regression models will often help us deal with more general models for characterizing the imprecision and inaccuracy of measuring instruments, which we discussed in Chapter 2. To illustrate, let us return to the basic models as given in Eq. 2-15, in which we accounted for instrumental biases and random errors of measurement. In doing so we estimated the imprecision of measurement as the standard deviation of the random errors of measurement, and we could also estimate the difference in constant biases as indicated, for example, with Eq. 2-19. However, suppose there is some trend in biases or systematic errors of

the instruments with the level of the characteristic measured or another parameter. What can be done concerning an appropriate analysis for such cases? It is very instructive to illustrate this with an example taken from the 1979 International Ozone Rocket Sonde Intercomparison (IORI) study. We acknowledge our appreciation to the Federal Aviation Administration (FAA), the National Aeronautics and Space Administration (NASA), and the World Meteorological Organization for the use of the data on ozone measurements in the stratosphere with instruments aboard rocket firings. Further studies of these data are underway.

In this example we will focus on the problem of determining the relative precision and accuracy of instruments for determining the ozone concentration in the stratosphere. Originally, it was desired to apply the three-instrument case of Chapter 2 by mounting three instruments aboard a rocket to take simultaneous measurements of stratospheric ozone concentrations as a function of altitude during flight of the rocket. However, this particular part of the overall experiment involves only one instrument on each of three rockets that were actually fired about an hour apart. In view of this, the most direct measure of the difference in errors of measurement for any two of the instruments for a given level of ozone concentration is not available although the principle and importance of using three instruments to study imprecision and inaccuracy of measurement may still be illustrated. Furthermore, the results from more extensive analyses could be that no significant change in ozone structure occurred during the three rocket flights. It will be seen in this connection that the imprecision of measurement varies with the altitude (and hence ozone concentration) and also that the differences in biases or systematic errors between pairs of instruments follow a trend with altitude. This example is, therefore, a rather general account of the basic principles of Chapter 2 on errors of measurement, precision, and accuracy of measurement along with the use of least squares fits of data covered in Chapter 6.

Although the reader may note some repetition of the basic principles outlined in Chapter 2, we believe, nevertheless, that a full account of the three-instrument approach to the analysis of ozone concentrations including the models of constant biases and variable biases will make our example more useful to the reader who may have very similar applications.

6A-2 ACCOUNT OF THE INTERNATIONAL OZONE ROCKET SONDE INTERCOMPARISON (IORI) STATISTICAL ANALYSIS

6A-2.1 THE THREE-INSTRUMENT APPROACH (CONSTANT BIASES)

The primary purpose of the statistical analysis was to determine the precision and accuracy of each instrument used in sampling the atmosphere; this would also give a comparison of the capabilities of the various types of instruments. First, however, we must define the terms precision and accuracy, which stem from errors of measurement introduced in making observations.

By precision we mean a suitable measurement of the variation in the errors of measurement of an instrument over a series of observations that are made with the instrument. Thus if this variation is "small", then the instrument is said to be "precise", but the larger the variation is the more imprecise the instrument and its measurements. Hence an estimate of the standard deviation of the errors of measurement of the instrument could be called the "imprecision" of measurement, and we will therefore estimate the imprecision of measurement by using the standard deviation of the errors of measurement to describe it. The estimation of the standard error of measurement is not very straightforward, however, because the observation or measurement taken consists of inseparable components, namely, the true value of the quantity measured, plus the bias or systematic error of the instrument used in the measurement process, plus a randomly varying error of measurement of the instrument. The problem then is to find a method of determining, i.e., stripping out, the standard deviation of the errors of measurement of each instrument by using a components of variance analysis. It is easily seen that if two instruments are used to take measurements on the same series of items or characteristics, the difference in the readings of the two instruments renders the difference in the random errors of measurement of the two instruments plus their difference in biases, or "calibration" values, so to speak. The variance of the series of differences will clearly give the sum of variances of the random errors of measurement of the two instruments since we might well assume that the biases of the instruments do not vary appreciably over a relatively short series of measurements—perhaps! We see, nevertheless, that even for two instruments taking the same series of measurements, the result is an estimate of the *sum* of the variances in the

random errors of measurement of the two instruments, and these are not yet separable. Hence we must continue in our analysis, especially if the function we are studying, such as ozone concentration versus altitude, varies considerably over the range of altitudes of interest. In fact, it becomes important to note that if three instruments are used to take the same series of measurements, we have three sets of differences in the random errors of measurement of the three instruments and their three differences in biases. However, when we find the variances of the three sets of differences in the errors of measurement, the result is simply three equations and three unknowns, which can easily be solved for the variances in the errors of measurement for each of the three instruments. The square roots of these final numbers give the standard errors of measurement of the three instruments or the three "imprecisions", except for the varying biases or systematic errors of the instruments, which may come into some prominence as indicated in the sequel. Varying biases may well exhibit a trend.

Clearly, the standard errors of measurement for each instrument, or the "imprecisions", are required to determine whether the mean biases are significant and hence can be estimated in size. Note in this connection that the average difference in the readings of any two instruments making the same measurements—when multiplied by the square root of the number of differences and then divided by the estimate of standard deviation of the differences based on $(n-1)$ degrees of freedom—can be used as an ordinary Student's t test to determine whether the two instrumental biases are significantly different in size. If no significance appears, one may conclude that the two instruments have equal biases (or read the same) although both may be nonzero.

Summarizing at this point, we see that the use of three instruments to take the same series of measurements will lead to a very desirable state of affairs, namely, a complete separation of the errors of measurement from the true values of the quantities we are attempting to measure, and this condition leads to a simple means of estimating the imprecisions or components of variance. If the analysis is straightforward, one may expect to determine estimates of the imprecisions of measurement of the individual instruments. There could be some complications, however. Those investigators who have worked with component of variance analyses know that often they encounter negative estimates of variance—which is disturbing, to say the least! These negative estimates of variance may be due to sampling, i.e., the vagaries of small sample size, or they may be due to the existence of "outliers" that have crept into the data and do not really represent the true characteristics of the instrumentation. The investigator may sometimes be able to decide to "throw out" anomalous values based on sound physical reasoning. However, most often he will not be able to make any such judgment, and some kind of statistical procedure for screening the data becomes quite necessary. There is a large body of statistical literature on the subject of detecting outlying observations in samples, such as Chapter 3, which may be resorted to as necessary. Alternatively, it is sometimes informative and satisfactory to ignore outliers and otherwise penalize precision of measurement by leaving them in the data. Once the "true" outliers have been screened, one may proceed to use the technique referenced in Chapter 3. For the vagaries due to sample size, usually it will be necessary to continue accumulating data until more stable estimates are available. Finally, we must remark that the model may not be sufficiently accurate to fit the data. These are, unfortunately, some of the pitfalls that may often enter the analysis. A rather full treatment of these topics along with optimum statistical techniques for estimating imprecision of measurement when two or more instruments are used is given in detail in Chapters 2 and 6 and Refs. 1 and 2.

As an allied check on the previously described procedure, and especially in view of the negative estimates of variance, one might well consider the approach that follows. Suppose a given measuring instrument or technique is used to determine the ozone concentration in the upper atmosphere. If the instrumental measurements show small scatter about some fitted curve, they may be said to be precise. Nevertheless, the instruments could have a constant *or* variable bias. In any event, the scatter about the curve, or the "residual variance", which is a measure of instrument imprecision, may be determined—even though the exact shape or form of the curve is unknown—by the methods of Ref. 3. The residual dispersion, so estimated for each instrument, also may be used as an estimate of imprecision although it may often include a bit more than the variance (or standard deviation) of just the errors of measurement. Nevertheless, for the case of the three instruments previously described, the estimate of the total variance in errors of measurement $(\hat{\sigma}^2_{e_1} + \hat{\sigma}^2_{e_2} + \hat{\sigma}^2_{e_3})$ (which is positive) can be scaled proportionately according to the size of the three residual variances,

hopefully, to give reasonably practical estimates of imprecision. Another way to estimate the residual dispersion about a curve for each measuring instrument would be to use least squares (orthogonal polynomials since the data are taken at equally spaced heights) and determine the residual variance. See, for example, any standard statistical textbook or Chapter 6. The residual variance for each instrument so determined is always positive and also measures imprecision.

The analysis outlined here determines the average imprecision of measurement of each instrument, which is that value "near the middle" of the data or measurements. However, since there are three instrument readings of ozone for each altitude, one may fit a least squares line or curve through the residual variances or standard deviations of the three instrumental measurements for each and all the altitudes to observe just how the standard error of measurement scales with height. For example, the standard deviations at various altitudes may increase or decrease with altitude and, hence, are so emphasized here for further reference.

As previously stated, the average difference in readings of any two instruments gives an estimate of the difference in (constant) biases. (A changing bias is treated in par. 6A-2.2.) Bias and imprecision together determine total inaccuracy.

The statistical model to which we have referred so far is of the general form:

$$O_{ij} = \omega_i + \beta_j + e_{ij}, \qquad i = 1,2,\ldots,n; \quad j = 1,2,3 \tag{6A-1}$$

where

O_{ij} = observed ozone concentration at altitude h_i for instrument j

ω_i = $\omega(h_i) = f(h_i)$ true but unknown ozone concentration at h_i, which varies with altitude as indicated

β_j = constant bias or systematic error of instrument j over the altitudes h_i of interest (Trends are also considered—see par. 6A-2.2.)*

e_{ij} = random error of measurement of instrument ($j = 1,2,3$) at height h_i, and $e_{ij} = N(0,\sigma_{e_j})$, i.e., e_{ij} is normally distributed with zero mean and imprecision of measurement σ_{e_j} for instrument j.

(Note: Compare Eq. 6A-1 with Eq. 2-15.)

By these definitions of terms, we see that the instrument with the smallest σ_e, or standard error of measurement, is the more precise one, and β determines the size of its bias or systematic error if it is significantly different from zero. If an instrument possesses good precision of measurement, i.e., σ_e is small, its bias relative to a standard or reference instrument may be detected and the instrument "recalibrated" to improve accuracy. (It may be difficult to reduce σ_e and thereby make the instrument more precise!) In any event, and with this description, one should begin to understand the meanings of "precision" and "accuracy" as applied here. Note that we have preferred to keep the imprecision σ_e and the bias β separate; for with the estimate of bias $\hat{\beta}$ and imprecision $\hat{\sigma}_e$ tagged onto each instrument, we know the capabilities of that measuring device. σ_e refers to the standard error of measurement of a single observation made and, hence, not an average value.

As a preliminary mode of orientation and an example of the three-instrument, constant bias assumption case, consider an analysis based on the mixing ratio** on the parts per million (ppm) scale. Suppose for example, that we obtained the following estimates of imprecisions and differences in biases obtained over 27 altitudes:

$$\hat{\sigma}_{e_1} = 0.30,\ \hat{\sigma}_{e_2} = 0.30, \text{ and } \hat{\sigma}_{e_3} = 0.10$$

$$\hat{\beta}_1 - \hat{\beta}_2 = 0.01,\ \hat{\beta}_1 - \hat{\beta}_3 = 0.20, \text{ and } \hat{\beta}_2 - \hat{\beta}_3 = 0.19 \text{ with } n = 27 \text{ altitudes.}$$

These estimates are easily found from Chapter 2. Here, we have taken instrument 1 as a "reference" instrument.

*We have noted that even the β_j vary with altitude.

**The term "mixing ratio" means the number of molecules of ozone per cubic centimeter of the sample divided by the number of molecules of air in that same volume, expressed in ppm.

We note that instruments 1 and 2 are equally precise, with equal imprecisions $\hat{\sigma}_{e_1} = \hat{\sigma}_{e_2} = 0.30$, but that instrument 3 is perhaps much more precise than 1 or 2. Whether the last statement is true may be determined from a significance test as in Eq. 2-78.

Since the biases of instruments 1 and 2 are also very nearly equal, it can be said that these two instruments are both equally precise and equally accurate.

Now consider instrument 3, which presumably is more precise than instruments 1 and 2, but reads 0.2 below instruments 1 and 2—due to calibration, perhaps. We may first determine whether the difference in biases of instruments 1 and 3 is significantly different from zero. This is easily accomplished with a t-type test similar to Eq. 2-63, i.e.,

$$t = (0.20 - 0)\sqrt{27}/[(0.30)^2 + (0.10)^2]^{1/2} = 3.3 .$$

Thus the difference in biases is real for 26 df.

Even though instrument 3 is more precise than either instrument 1 or 2, it may not be more accurate; this depends on how many observations may be made with instrument 1 or 2 (as compared with instrument 3) and "averaged", for example. However, if instrument 3 is recalibrated to eliminate the bias of 0.20, instrument 3 becomes more precise and more accurate than instrument 1 or 2—assuming the reference instrument was properly calibrated.

Finally, if the three instruments were of the same design and similarly produced, with about equal precision, the average imprecision of measurement for such an instrument may be estimated from

$$\hat{\sigma}_{av} = [\hat{\sigma}^2_{e_1} + \hat{\sigma}^2_{e_2} + \hat{\sigma}^2_{e_3})/3]^{1/2} \qquad \text{(6A-2)}$$

and this is also the square root of one-sixth of the sum of the three variances of the differences in errors of measurement of the three instruments taken two at a time. For the same type of instrument, it would seem that this quantity could be taken as the average imprecision of measurement. For the given data one would find that the average standard error of measurement for an instrument of this type would be about

$$\hat{\sigma}_e = 0.25$$

although if $\hat{\sigma}_{e_3}$ is significantly lower than the estimated standard errors of measurement of instruments 1 and 2, no such averaging should be encouraged.

Recall that the imprecisions σ_e may need to be scaled with altitude or with the amount (level) of ozone in the atmosphere and that this could be investigated separately from this particular analysis. Such scaling of the σ_e's may be done with a least squares fit on the estimated standard deviations at each or several of the altitudes, for example. However, if the instrumental biases vary with altitude, as we discuss in the sequel, some care has to be exercised to assure that one is working with residual deviations of a random, nonsystematic character.

Finally, since we have detected a significant average bias for instrument 3, steps should be taken to make appropriate adjustment or to recalibrate the instrument. In fact, recalibration of the instrument might involve a bias correction that varies with altitude as a "trend", if such is the case.

Although some of the IORI data may appear to be well represented by the simple model just discussed, involving a fairly constant bias along perhaps with some scaling of σ_e, there also appears to be significant drifting of measurements between instrumental reading pairs. Therefore, we will consider this type of problem next, especially for biases changing with altitude or level of ozone because these will also have an impact on the residual dispersion. There seems to be little point, however, in adopting a complex model for analytical purposes when a simpler one will suffice. On the other hand, we should be on the lookout for either a changing σ_e or for any drifts in evident instrumental biases as a function of either altitude or level of ozone measured. Any such changes often are found by simply plotting the σ_e and the $(\beta_i - \beta_j)$ (determined by differences between pairs of instrument readings) versus altitude.

6A-2.2 ESTIMATION WHEN INSTRUMENTAL BIASES CHANGE WITH ALTITUDE OR OZONE LEVEL

Although the simple model of Eq. 6A-1 is based on the assumption of a constant bias or systematic error for the instruments, it could be extended to a more complex one, or "generalized". However, it is to be expected that the analysis would become more complex and perhaps cost more by employing additional instruments. Nevertheless, we have made some preliminary plots and find that the difference in instrument readings shows a relation with either the altitude or the level of ozone measured.

In the original 1948 study by Grubbs (Ref. 1), the ω_i represented a random variable, i.e., the running times of fuzes, and the biases were evidently small. If there existed a *linear* relation between the systematic error of measurement or bias and the level of ozone, then the model of Jaech (Refs. 4 and 5) developed in 1964 might well apply. Jaech's model is expressed as

$$O_{ij} = \alpha_j \omega_i + \beta_j + e_{ij} \tag{6A-3}$$

where now the jth instrument scales the true ozone level ω_i with a slope α_j (somewhat near unity). Note that when $\alpha_j = 1$, the model of Eq. 6A-3 is exactly the same as that of Eq. 6A-1. The systematic error, now consisting of the first two terms of Eq. 6A-3, also becomes much more involved due to the varying instrumental biases. For example and in view of Eq. 6A-3, the difference in biases for instruments 1 and 2 now becomes

$$(\alpha_i \omega_i + \beta_1) - (\alpha_2 \omega_i + \beta_2) = \beta_1 - \beta_2 + (\alpha_1 - \alpha_2)\omega_i \tag{6A-4}$$

which is linear in the amount of ozone present. Jaech's analysis evolved in connection with a study of reactor fuel element quality (Refs. 4 and 5), for which the assumption of Eq. 6A-3 appeared reasonable. Moreover, there is little difficulty in estimating the imprecisions σ_{e_j}, the difference in the β_j, or the difference in the additional coefficients α_j. The reader may study Refs. 4 and 5 for details.

In a very similar way, a linear systematic error model may be set up and used by substituting a function of the altitude h_i in place of the ozone concentration ω in Eq. 6A-3. However, neither of these two linear models is ample to satisfy the requirements arising here. One should appreciate this position by referring to the data of Table 6A-1, which we will use to conduct a typical analysis. The data represent measurements of ozone from the three Kreuger instruments (UV absorption) to determine ozone amounts on Super Loci Rocket Flights 249, 250, and 251, which were fired about 45 min before noon, at noon, and about 45 min after noon, respectively. Thus neither of the three instruments is on the same vechicle, nor do the instruments determine ozone amounts simultaneously. Thus one might suspect differences between instrument readings due to a variety of causes. One of the very striking occurrences is that between 25 and 50 km the concentration of ozone varies about one hundred to one in some systematic way, and it is far from linear! Recall from Eq. 6A-1 or Eq. 6A-3, that we need an estimate of the (random) residual dispersion to determine the imprecision. The last three columns (Table 6A-1) of differences between readings of pairs of instruments show rather severe trends or raggedness, very much nonlinear. Thus, we have had to decide against the use of models, such as Eqs. 6A-1 and 6A-3, but in favor of a very significant extension of them. We again start in a similar manner with the variances of the differences, or really sums of squares (SS), and delete components that arise as a result of the trends in biases with altitude.

Before proceeding with the suggested analysis, a remark or two concerning transformations of the original data is in order. Some other scales of analysis have been suggested, and consideration has been given to them. They include the mixing ratio (or number of ozone molecules divided by the number of air molecules in a cubic centimeter), the normalized number density (or observed ozone density divided by the Kreuger-Minzer (Ref. 6) standard values at each altitude), and an analysis based on logarithms of the original ozone measurements. The mixing ratio and the normalized number density both involve scaling numbers that are different at each altitude and, hence, are nonuniform transformations. Thus conversion from the scale of analysis back to the original ozone data becomes very difficult. The use of the logarithmic transformation seems to work quite well and even reduces or smooths out the effect of some outlying values when transformed back to the original data. However, an analysis on the logarithmic scale does not appear to reduce the need for higher order fits on

TABLE 6A-1*

ORIGINALLY MEASURED CONCENTRATIONS OF OZONE IN NUMBER OF MOLECULES PER CUBIC CENTIMETER AND INSTRUMENT DIFFERENCES FOR THE THREE KREUGER INSTRUMENTS (UV ABSORPTION) ON FLIGHTS 249, 250, AND 251

Altitude, km	Inst 1	Inst 2	Inst 3	I1-I2	I2-I3	I3-I1
	(Original ozone concentrations divided by 10^{11})					
25	45.6	39.9	40.8	5.70	-0.90	-4.80
26	42.8	38.4	41.9	4.40	-3.50	-0.90
27	39.7	35.9	40.1	3.80	-4.20	0.40
28	36.8	36.1	36.9	0.70	-0.80	0.10
29	35.3	32.7	34.5	2.60	-1.80	-0.80
30	33.1	30.9	32.9	2.20	-2.00	-0.20
31	31.2	30.9	30.9	0.30	0.00	-0.30
32	27.9	25.9	26.5	2.00	-0.60	-1.40
33	23.5	21.8	22.3	1.70	-0.50	-1.20
34	20.8	19.8	20.0	1.00	-0.20	-0.80
35	18.0	16.2	16.8	1.80	-0.60	-1.20
36	14.4	13.4	14.4	1.00	-1.00	0.00
37	11.9	11.8	12.2	0.10	-0.40	0.30
38	10.1	9.96	10.0	0.14	-0.04	-0.10
39	8.14	8.26	8.12	-0.12	0.14	-0.02
40	6.50	6.37	6.71	0.13	-0.34	0.21
41	5.45	5.31	5.53	0.14	-0.22	0.08
42	4.62	4.50	4.48	0.12	0.02	-0.14
43	3.56	3.39	3.46	0.17	-0.07	-0.10
44	2.82	2.57	2.60	0.25	-0.03	-0.22
45	2.01	2.08	1.99	-0.07	0.09	-0.02
46	1.55	1.59	1.65	-0.04	-0.06	0.10
47	1.31	1.19	1.26	0.12	-0.07	-0.05
48	0.877	0.930	0.956	-0.053	-0.026	0.079
49	0.550	0.707	0.740	-0.157	-0.033	0.190
50	0.480	0.525	0.605	-0.045	-0.080	0.125

the transformed scale; thus one may as well deal with the original ozone measurements. It is for these reasons that our analysis is directly on the original measurements in order to isolate random errors and systematic bias trends.

Since we are expressing the imprecision of measurement as the standard deviation of the errors of measurement—this should be about equivalent to the residual dispersion remaining after meaningful trends have been eliminated—two preliminary procedures suggest themselves. First, we may apply the technique of Morse and Grubbs (Ref. 3) to obtain a stable estimate of residual dispersion by working with higher order differences for the readings of an instrument with increasing altitude. A positive estimate of the residual standard deviation always results from such analysis. Secondly, since the ozone concentrations are listed for equally spaced altitudes, we may use orthogonal polynomials to fit either a line, a parabola, a cubic, etc., and terminate at an insignificant fit. The residual dispersion remaining about the fitted curve then could be taken

*Some further refinements in data reduction have altered these data somewhat, but the illustrative value remains.
These preliminary instrument readings were obtained with permission from Dr. Arlin Kreuger of NASA, Greenbelt, MD.

as an initial estimate of the imprecision for that instrument. Of course, we would work finally with the differences in measurements of two instruments at a time to estimate the standard errors of measurement for each instrument.

Referring now to the second column of Table 6A-1 and the measurements of Kreuger instrument 1 on Flight 249, we found using orthogonal polynomials that linear, quadratic, cubic, and quartic regressions are all highly significant with a residual variance of about 0.54×10^{22} from the quartic fit and about 0.49×10^{22} from the insignificant quintic fit. Thus it is seen that the standard deviation expressing imprecision should be about 0.7×10^{11} mol/cm^3. We have used instrument 1 only as an illustration although it would be highly desirable to know the "best" (more precise and accurate) instrument and to use it as a reference or "standard".

Having arrived at an indication of the approximate imprecision of measurement, we now turn to an analysis of the difference in readings of two instruments since that difference should reflect only the difference in errors of measurement of the two instruments and also show trends in systematic errors as a function of altitude. In order to examine this, we will analyze the difference in ozone determinations of instrument 1 and instrument 2, i.e., the fifth column of Table 6A-1, which indicates a rather severe trend for instrumental bias differences ranging from large positive differences at 25 km to small negative differences at 50 km. Therefore, at the lower altitudes instrument 1 gives readings much higher than instrument 2. Some of this difference could perhaps be due to a change in ozone levels within the 45-min lapse time although it could well be instrumental differences arising from calibration problems *et al.* By taking the SS of the figures (differences listed) in column 5, Table 6A-1, about their mean, the result is 60.93×10^{22}, which, when divided by 25 df, estimates a variance of 2.437×10^{22}, or standard deviation of 1.56×10^{11}. Such values are noticeably larger than perhaps expected as a measure of the dispersion of differences in errors of measurement. Consequently, we should look for a trend in the instrumental bias differences of instrument 1 and instrument 2. Perhaps we could fit a line or higher degree curve to these instrumental bias differences as a function of the altitude h. Such an analysis has been carried out and is indicated on Table 6A-2, where we have used orthogonal polynomials in the process of fitting a line, a parabola, or a cubic equation. Note that data for the $n = 26$ altitudes have been reduced to 13 pairs in the form of sums s_i or differences d_i since only half of the orthogonal polynomial values are listed for n greater than about 8. (See the last three columns at the top of Table 6A-2.*) The sums s_i for each pair of altitudes are to be multiplied by the ξ's with even subscripts (powers), and the differences d_i are to be multiplied by the ξ's with odd orthogonal polynomial powers as in Table 6-6.

The fitted equation is of the form (for O = ozone):

$$\left.\begin{aligned} O &= a + bh + ch^2 + dh^3 + \cdots \\ &= A + B\xi_1 + C\xi_2 + D\xi_3 \\ &= \bar{O} + B'\xi_1' + C'\xi_2' + D'\xi_3' + \cdots \end{aligned}\right\} \quad \text{(6A-5)}$$

where

$$A = \bar{O} = 1.0725 = \text{constant term} \quad \text{(6A-6)}$$

$$\xi_o = 1$$

$$\xi_1 = h_i - \bar{h}_i = h_i - 37.5 \quad \text{(6A-7)}$$

$$\xi_{r+1} = \xi_1\xi_r - r^2(n^2 - r^2)\xi_{r-1}/[4(4r^2 - 1)] \quad \text{(6A-8)}$$

$$\xi_r' = \lambda\xi_r \quad \text{(6A-9)}$$

*Reread also the paragraph just above Eq. 6-132 in Chapter 6.

where

h_i = ith altitude

$\bar{h}$ = average altitude = 37.5.

TABLE 6A-2

ANALYSIS OF DIFFERENCE IN BIASES BETWEEN INSTRUMENTS 1 AND 2 VERSUS ALTITUDE

Paired Altitudes h_i, km	Sum for Altitudes s_i	Difference for Altitudes d_i	(From Table 6-6 for $n = 26$) ξ'_1	ξ'_2	ξ'_3
37,38	0.24	0.04	1	-28	-84
36,39	0.88	-1.12	3	-27	-247
35,40	1.93	-1.67	5	-25	-395
34,41	1.14	-0.86	7	-22	-518
33,42	1.82	-1.58	9	-18	-606
32,43	2.17	-1.83	11	-13	-649
31,44	0.55	-0.05	13	-7	-637
30,45	2.13	-2.27	15	0	-560
29,46	2.56	-2.64	17	8	-408
28,47	0.82	-0.58	19	17	-171
27,48	3.75	-3.85	21	27	161
26,49	4.24	-4.56	23	38	598
25,50	5.65	-5.75	25	50	1150
From Table 6A-1:	$\bar{O} = 1.0725$	Divisors:	5850	16,380	7,803,900
		Coefs: λ =	2	1/2	5/3

Divisors are the sums of squares, $\Sigma(\xi'_i)^2$
Constant term = $\bar{O} = 1.0725$
Coefficient of linear term = $\Sigma d_i \xi'_i / \Sigma(\xi'_1)^2$
Sum of squares for linear regression = $(\Sigma d_i \xi'_1)^2 / \Sigma(\xi'_1)^2$, etc.

Part of Table 6A-2 is taken from Table XXIII of Fisher & Yates: *Statistical Tables for Biological, Agricultural, and Medical Research,* published by Longman Group Ltd., London (1974) 6th edition (previously published by Oliver & Boyd Ltd., Edinburgh) and by permission of the authors and publishers.

TABLE 6A-3

ANOVA OF TRENDS IN DIFFERENCES (COLUMN 5, TABLE 6A-1) OF BIASES, INSTRUMENT 1 MINUS INSTRUMENT 2

Source of Variation	SS	df	Residual SS	df	Residual Variance	F Ratio
Total	60.93	25				
Linear Regression	38.10	1	22.83	24	0.951	Highly Sig.
Quadratic Regression	10.30	1	12.53	23	0.545	Highly Sig.
Cubic Regression	2.01	1	10.52	22	0.478	Not Sig.

The ξ' are always in integral values; they are made so by the proper choice of the λ's listed on Table 6A-2.

The coefficients B', C', and D' are determined from

$$B' = \Sigma\xi_1' d_i / \Sigma(\xi_1')^2 = -0.0807 \tag{6A-10}$$

$$C' = \Sigma\xi_2' s_i / \Sigma(\xi_2')^2 = +0.0251 \tag{6A-11}$$

$$D' = \Sigma\xi_3' d_i / \Sigma(\xi_3')^2 = -0.000507, \text{ etc.} \tag{6A-12}$$

Finally, the SS for linear regression, quadratic regression, cubic regression, etc., are found simply by multiplying the appropriate estimated coefficients—B', C', D', etc.—again* by the numerators of Eqs. 6A-10, -11, and -12. The SS values are brought together in our ANOVA Table 6A-3 of bias difference trends. By inspection of this part of the table, one notes that the fit of the cubic regression is not significant statistically; thus we would terminate at a fit of the quadratic equation or parabola. This would mean that the final fit to the differences in biases or systematic errors of instrument 1 and instrument 2 would be—by Eqs. 6A-6 through 6A-9—with $n = 26$

$$\delta O_{1-2} = \hat{\beta}_1 - \hat{\beta}_2 = 1.0725 - 0.0807\xi_1' + 0.0251\xi_2' \tag{6A-13}$$

where

$$\xi_1' = 2(h_i - 37.5) \tag{6A-14}$$

$$\xi_2' = [(h_i - 37.5)^2 - 56.25]/2. \tag{6A-15}$$

Eqs. 6A-14 and 6A-15 may be substituted as desired into Eq. 6A-13 to obtain the direct relation between the trend of the difference in systematic errors of instrument 1 and instrument 2 as a function of the altitude h. This result expresses the difference in calibrations or biases of instruments 1 and 2 over the range of altitudes, 25 km to 50 km, and clearly represents a very significant trend. Further calibration of instruments 1 and 2 may be obtained by reference to an appropriate standard. In summary, we say that the bias errors are not constant and thus introduce a significant problem indeed. Also of importance to us is the estimate of imprecision, which may be determined by using the residual variance resulting from the quadratic fit. Thus it is seen from the next to bottom line of Table 6A-3 that the residual variance about the quadratic fit is 0.545, which is a measure of the variance of the difference in unaccounted-for errors of measurement between instruments 1 and 2 or, that is, the sum of variance in errors of instrument 1 and the variance in errors of instrument 2. The value 0.545, therefore, is a more appropriate value to use in the method of Grubbs (Refs. 1 and 2) for estimating the imprecisions of measurement. This residual variance of 0.545 will be used after we have made similar determinations for instruments 2 and 3 and instruments 3 and 1 in order to model the three-instrument case.

Return to Table 6A-1 and the sixth column of differences for the determination of ozone by instruments 2 and 3. An analysis similar to that carried out for the differences of instruments 1 and 2 leads to the quadratic fit

$$\delta O_{2-3} = \hat{\beta}_2 - \hat{\beta}_3 = -0.6623 + 0.0474\xi_1' - 0.0147\xi_2' \tag{6A-16}$$

with ξ_1' and ξ_2' the same as in Eqs. 6A-14 and 6A-15 and a residual variance based on 23 df of 0.5575.

An analysis of the differences between the ozone determinations of instruments 3 and 1 leads to a significant linear fit only, which is

$$\delta O_{3-1} = \hat{\beta}_3 - \hat{\beta}_1 = -0.4102 + 0.0328\xi_1' \,, \tag{6A-17}$$

*That is, SS for linear regression is

$$\text{SS} = B'(\Sigma\xi' d_i) = \frac{(\Sigma\xi_1' d_i)^2}{\Sigma(\xi_1' d_i)^2}.$$

with a residual variance now based on 24 df equal to 0.823.

With reference to the systematic trends of differences in biases between instruments, the larger slope of the quadratic terms is the coefficient of 0.0251 in Eq. 6A-13 involving the first instrument although there appear to be some calibration problems for all three of the instruments unless it is known which, if any, is correct!

For estimates of the imprecisions of measurements for the three instruments near the middle of the range, or central altitudes, the three residual variances about the statistically significant fits may now be used in Grubbs' methodology (Refs. 1 and 2). In fact, immediately after eliminating significant trends, we have three equations and three unknowns, i.e.,

$$\hat{\sigma}_{e_1}^2 + \hat{\sigma}_{e_2}^2 = 0.545$$

$$\hat{\sigma}_{e_2}^2 + \hat{\sigma}_{e_3}^2 = 0.558$$

$$\hat{\sigma}_{e_3}^2 + \hat{\sigma}_{e_1}^2 = 0.823$$

or, solving for the three unknowns,

$$\hat{\sigma}_{e_1}^2 = 0.405, \quad \hat{\sigma}_{e_1} = 0.636 \times 10^{11} \text{ mol/cm}^3$$

$$\hat{\sigma}_{e_2}^2 = 0.140, \quad \hat{\sigma}_{e_2} = 0.374 \times 10^{11} \text{ mol/cm}^3$$

$$\hat{\sigma}_{e_3}^2 = 0.418, \quad \hat{\sigma}_{e_3} = 0.647 \times 10^{11} \text{ mol/cm}^3.$$

We note first that the estimate $\hat{\sigma}_{e_1}$ of 0.64 is a bit smaller than the value 0.70 left as a residual sigma had we fit a quintic to the data or readings of the first instrument. This provides somewhat of a check.

Of course, it may be that the σ_{e_j} actually increase in value toward the lower altitudes and are smaller for the higher altitudes. Such scaling might be estimated by using the standard deviations of a number of residuals at each end of the fitted curves.* However, what seems to be of much concern are the trends in the differences in bias or systematic errors between pairs of instruments. For example, for instruments 1 and 2 the estimated difference in biases at $h = 25$ km is from Eqs. 6A-13 through 6A-15

$$\delta O_{1\text{-}2} = 1.0725 - 0.0807(-25) + 0.0251(50) = \hat{\beta}_1 - \hat{\beta}_2 = 4.35$$

with a residual of

$$5.7 - 4.35 = 1.35 \text{ (still unaccounted for)}$$

versus a $\hat{\sigma}_{e_1 - e_2}$ of about $(0.545)^{1/2} = 0.74$ (average, unexplained).

The same type of analysis outlined here may also be used for the other instruments involved in the intercomparison study.

There certainly needs to be a tie-in between the difference in calibration curves (bias trends) of the three Kreuger instruments analyzed here and all of the various types of instruments from other countries (Australia, Canada, India, Japan, and U.S.). Some standard may be needed here. For the Nike Orion triad flights, some very valuable comparisons may be made since the Australian, Canadian, Indian, and Japanese instruments were aboard the same rocket flights.

It is clear that with a good reference profile the bias trends of the instruments could be removed completely!

In summary, one observes from Eqs. 6A-1, 6A-3, and models—such as Eqs. 6A-13, -16, and -17 for systematic error differences or instrument calibration problems—that the true ozone concentration ω varies

*There is some evidence from residuals that the σ_e near 25 km may be several times that at 50 km—the data are rough and limited. The average σ_e for the three instruments at 25, 37, and 50 km are estimated to be about 1.1, 0.61, and 0.14×10^{11} mol/cm^3, respectively.

perhaps in a complex manner with altitude. To study the precision and accuracy of measurement, however, we must work with differences in the readings of the instruments taken two at a time to eliminate the amount of ozone present so that random errors of measurement or imprecision on one hand and the differences in instrumental biases on the other may be modeled and estimated. In fact, the systematic errors β_j vary with altitude either in a significant quadratic or linear manner, giving rise to statistically described bias trends expressed as systematic error differences of the instruments. Once the significant bias trends are determined, the residual scatter may be used to estimate the average imprecision of measurement σ_{e_j} of the instruments.

For the overall accuracy problem, it can be said that one first experiences a variable bias in the instrumental readings as expressed in Eqs. 6A-13, 6A-16, or 6A-17. Depending on which particular instrument, if any, is actually correct, one is unsure just what the true calibration curve of each instrument is or should be. Once such systematic errors are incurred, one should expect that the random errors of measurement of the instruments may vary with altitude and be described by a standard error of perhaps about 1.1×10^{11} mol/cm^3 at 25 km to about 0.14×10^{11} mol/cm^3 at 50 km. With such a varying imprecision of measurement depending on the altitude, it becomes clear that once the trends of the differences in biases between instruments have been eliminated from the original differences, then for each altitude one can determine the three residual differences and the average of these three differences for each altitude. This average difference at each altitude could then be divided by $2 \times 0.5642 = 1.1284$ to give an estimate of the imprecision sigma at that altitude. Finally, a least squares fit on these estimates with altitude will give the estimated imprecision of measurement curve for the three instruments of the same type represented.

In summary then, there is an instrumental bias difference curve between instruments of a type, for each instrument apparently has its own bias trend, and for the standard deviation of the imprecision of measurement, there is also a fitted least squares curve or trend representing the average value of the three instruments of a particular type.

6A-3 GENERAL COMMENT ON LINEAR REGRESSION WITH ERRORS IN BOTH VARIABLES

In the example of this appendix, we saw that even though the differences in biases may follow a trend and the imprecision of measurement may vary with the level of the quantity of interest, one could, with the use of three instruments, model the situation with rather good accuracy. Because of such attainment, one is led to a reconsideration of the linear regression problem. Of wide interest is the case in which the true part of the dependent variable is a linear function of the true part of the independent variable and in which there are errors (of measurement) in both variables. It is well-known for this case that there are five basic parameters to be determined—i.e., the true slope, intercept, variance of the quantity of interest, and the variances of the errors in both the independent and dependent variables. However, these five parameters cannot be estimated satisfactorily without supportive ancillary information. Nevertheless, if it were possible for the linear regression problem to be treated as a three- or more instrument case with redundant measurements on either the dependent or the independent variable, sufficient overdetermination would be achieved so that the major parameters of interest could be estimated. One might well note again in this connection Eqs. 6-49 through 6-56 although we cannot go extensively into this area of statistical investigation here.

REFERENCES

1. Frank E. Grubbs, "On Estimating Precision of Measuring Instruments and Product Variability", Journal of the American Statistical Association **43**, 243-64 (June 1948).
2. Frank E. Grubbs, "Errors of Measurement, Precision, Accuracy and the Statistical Comparison of Measuring Instruments", Technometrics **15**, 53-66 (February 1973).
3. Anthony P. Morse and Frank E. Grubbs, "The Estimation of Dispersion from Differences", The Annals of Mathematical Statistics **18**, 194-214 (June 1947).
4. John L. Jaech, "A Program to Estimate Measurement Error in Nondestructive Evaluation of Reactor Fuel Element Quality", Technometrics **6**, 293-300 (August 1964).

5. John L. Jaech, "Estimating Within-Laboratory Variability From Interlaboratory Test Data", Journal of Quality Technology **11**, 185-91 (October 1979).
6. Arlin J. Kreuger and Raymond A. Minzner, "A Mid-Latitude Ozone Model for the 1976 US Standard Atmosphere", Journal of Geophysical Research **81**, 4477-81 (20 August 1976).

CHAPTER 7

ORDER STATISTICS AND APPLICATIONS

The use of sample order statistics by the Army analyst represents a very important and wide field of application that will continue to have prime demand. Indeed, the theory of sample order statistics is indispensable since many Army problems invariably result in truncated or censored samples for the observations taken. Order statistic theory is very useful in the following areas:

1. *Studies of the maximum dispersion or sample range*
2. *Mean values of order statistics as they relate to population parameters*
3. *Detection of outlying observations*
4. *Use of quasi-ranges when the sample extremes are suspect*
5. *Estimation of population parameters from truncated or censored samples*
6. *Use of simple, efficient linear estimators of the population mean and standard deviation*
7. *Statistics of extreme occurrences*
8. *Relationships to reliability and life testing problems*
9. *Analyses of the delivery accuracy of weapons including either rectangular coordinates or radial order statistics*
10. *Placing of confidence bounds on the proportion of the sampled population between limits*
11. *Determination of population characteristics from truncated target firings of weapons*
12. *Estimation of discrete population parameters, such as for the Poisson distribution.*

These topics are all discussed and presented in useful detail for the Army statistician or analyst, and several examples illustrating truncated sample theory are given.

7-0 LIST OF SYMBOLS

a_r = constant or coefficient related to rth sample order statistic, used especially in estimation of the mean from a linear form

$b = 1/\beta$ = reciprocal of shape parameter used in Eq. 7-27

b_r = constant or coefficient related to rth sample order statistic, used for estimation of the population sigma from a linear form

$E(\ )$ = used to denote expected or mean value of quantity in parentheses

$E(x_r^k)$ = kth moment about origin of rth order sample statistic

est = denotes estimate of parameter

$F(x)$ = cumulative probability distribution of random variable x

$F(u, v)$ = Snedecor-Fisher F statistic with u and v degrees of freedom (df)

$F_n(x) = Pr(x_n \leq x)$ = cumulative distribution of largest sample value x_n

$F_1(x)$ = cumulative distribution of smallest sample value x_1

$F_{1-\alpha}(u, v)$ = $(1 - \alpha)$ probability level of F, i.e., upper α probability level

$F^{-1}(\ )$ = inverse of function F

F' = ratio of two sample ranges

f = sum of frequencies for at least one hit

$f(x)$ = probability density function (pdf) of random variable x

f_x = observed number or frequency for x hits

f_0 = frequency for zero number of hits class

$G(\ )$ = cumulative distribution function of quantity in parentheses

g = number of rounds passing below a rectangular target

g, h = certain sums of the n_{ij} in Eqs. 7-61 and 7-62

h = number of rounds passing above a rectangular target

$I_x(u,v)$ = incomplete beta function ratio

$I_y(u,v)$ = Karl Pearson's incomplete beta function (see Eq. 7-7 or Ref. 7)

k = factor associated with tolerance limits

m = number of target misses

$m = r + s$ = total number of "blocks" or sample spaces below rth smallest sample order statistic and above the sth largest order statistic (see par. 7-7.5)

m_i and m_i' = Visnaw's notation for n_{ij} sums in Eqs. 7-59 through 7-62

m'' = number of rounds which cannot be determined as being left of, above, to the right of, or below the target

$N = f_0 + f$ = total frequency including zero class frequency

n = sample size

n_{ij} = number of target misses in the i, jth "quadrant" (see Eqs. 7-59 through 7-62)

$\binom{n}{r}$ = combination of n things taken r at a time

$P(c,n,p)$ = chance of occurrence of c or more successes in n trials when chance of occurrence in a single trial is p, i.e., the binomial sum

$P(h)$ = probability of h or more hits

$P_i(x)$ = failure-time distribution for ith component of a system

$Pr[v]$ = probability of event happening in v trials

p = number of dimensions—$p = 2$ for bivariate case

$p(x)$ = probability of exactly x hits

q = numerical quantity

$R(x) = 1 - F(x)$ = upper tail of distribution of x beyond the value x, and often referred to as the "reliability"

r = number of rounds passing to left of a rectangular target

$r_i = (x_i^2 + y_i^2)^{1/2}$ = ith radial order statistic about origin or center of impact

r_0 = cutoff radius for truncation of radial sample values

r_1 = number of smallest ordered sample observations censored in sample of n

r_1 = number of rounds missing target on left (Example 7-4)

r_2 = number of rounds missing target on right (Example 7-4)

r_2 = number of the largest ordered sample observations censored in sample of n

s = number of rounds passing to right of target

s = sample standard deviation, based on $(n - 1)$ degrees of freedom (df)

T = mean number of trials to an "occurrence", or between occurrences

$t = w/s$ = Studentized sample range

t_i = ith ordered time observation

t_r = time to rth failure

t_0 = specified truncation time

$u = \ln\theta$ = logarithmic transformation

Var( ) = denotes variance of quantity in parentheses

W = central area of a distribution between x_1 and x_n (see Eq. 7-34)

$w = x_n - x_1$ = sample range = w_0 also

$w_r = x_{n-r} - x_{r+1}$ = rth quasi-range of sample

$x(\alpha)$ = value of variable x directly related to the upper α probability level

x,y = rectangular coordinates of a point
$\bar{x}$ = sample mean
x_i = ith sample order statistic
x_i = ith ordered sample observation or value; we have that $x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_i \leq \cdots \leq x_n$
x'_j = jth sample observation in the order observations were taken
x_n = largest observation in sample
x_1 = smallest observation in sample
$Y_{[i]}$ = bivariate "concomitant" of ith sample order statistic X_i
β = confidence level
β = population parameter
β = shape parameter of a distribution
$\Gamma(\ )$ = complete gamma function of quantity in parentheses
γ = given fraction of the population
θ = mean value parameter for exponential distribution
θ = characteristic life, or a scale parameter
λ = parameter of a distribution = $1/\theta$ for exponential distribution and is the expected number of occurrences for the Poisson distribution in Eq. 7-63
μ = population mean
$\hat{\mu}$ = estimate of the parameter or mean value μ
μ^* = "optimal" estimate of μ, i.e., for example, a minimum variance estimate
$\sigma(\ )$ = standard deviation of the quantity in parentheses
$\hat{\sigma}$ = estimate of the population standard deviation or sigma
$\chi^2(\ )$ = random variable chi-square for number of degrees of freedom (df) given in parentheses
$\hat{}$ = denotes estimate of the quantity under it

7-1 INTRODUCTION

The last thirty years or so have witnessed an enormous growth in the applications of sample order statistics. This, no doubt, is due largely to the ever-increasing importance of life testing, reliability, availability, and maintainability of systems of all kinds, especially insofar as many Army applications are concerned. In addition, there are many practical applications for which the data naturally arise in order of magnitude, such as the life span in minutes, hours, days, and months of items or systems placed in service. Moreover, sometimes sample data are either truncated or censored, so that often one does not have available the smallest few or the largest few sample observations to analyze. Then again, it is also often true that the few largest and/or few smallest observations may not represent true sample values because they may be prone to shifts in level or other abnormal conditions. As a further example, one might consider a combat "experiment" for which he counts among the tanks knocked out exactly the number of hits scored by projectiles or antitank weapons from the other side. Note in this case that one can observe directly the number of tanks for which there is exactly one hit, the number of disabled tanks having two hits, etc., but he cannot take any direct observations on the number of times each of the other tanks in the battle was shot at, but not hit, so that truncation or censoring for this type of combat data occurs. The initial, total number of rounds fired in combat may, nevertheless, be of much importance either for a complete analysis or for logistical planning purposes. Thus we see some of the possible order statistic-type problems with which the analyst might be faced in some Army applications, including data censoring or truncation of some types.

First, we must define "order statistics" properly. We all are accustomed in sampling experiments to take or to have at hand some n observations, which ordinarily are listed in the order in which they were observed; namely, we have a "random sample of n". In the case of order statistics, however, the sample observations may even be observed in ascending order, such as for the lifetimes of items on test, or the sample values may be arranged in increasing order of magnitude of the measurements. To enforce some brevity of notation

throughout this chapter, we are specifying that the observations in the order in which they were originally observed are

$$x_1', x_2', x_3', \ldots, x_j', \ldots, x_n'$$

where we have used primes for the occurrence order. When the n sample values are placed in ascending order of magnitude, they become

$$x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_i \leq \cdots \leq x_n$$

where the quantity x_i is known as the ith order (sample) statistic. Generally, we will use the random variable x to describe the characteristic under study; however, it should be noted that often the physical characteristic of time is of much importance. For example, data on lifetimes or the failure times for an item or piece of equipment is the key variable studied. Hence we might well use the ith ordered time, or t_i, in place of x_i as an observation.

Very often we will find that the longer times-to-fail are not taken due to costly experimentation, for example, or it may be thought that only some r of n possible observations will be sufficient for the analysis purposes at hand. In such cases it is seen that only the first $r < n$ order sample statistics are available for analysis since the last or $(n - r)$ largest sample observations have been censored or truncated from the test or experiment. Sometimes the sample may be truncated or censored on the left side instead of on the right side.

For the entire sample of ordered observations, the highest observation x_n and/or the lowest observation x_1 may be of particular interest since they may be tested statistically by the method given in Chapter 3 to determine whether they are "outliers". Also it is well-known that the difference between the largest and the smallest sample values is the sample range (Chapter 3). Thus the sample range w is defined algebraically as

$$w = x_n - x_1. \tag{7-1}$$

We will discuss briefly the probability distributions of the smallest sample value, the largest sample observation, and the range in par. 7-2 because they are of use in many practical applications.

We proceed to present and discuss some of the many uses of order statistics and their important properties in connection with timely and unique analyses of experimental data. The Army applications of statistical methods involve many instances for which the analysis of ordered sample values is called for or even mandatory. Indeed, to cite another example, the weapon developer may have a new projectile under development, and he desires to estimate the round-to-round population standard deviation of the item. When test firings at a vertical target are carried out, however, some of the experimental projectiles may miss the target, so that the sample of rounds may be truncated above, below, to the left, and/or to the right of the target and only the coordinates of the impacting rounds are measurable. One immediately sees—as it actually turns out—that if the population mean and standard deviation can be estimated in an unbiased manner, the use of order statistic theory will be entirely justified. It is just such occurrences that often call for order statistic analyses—adding indispensable tools to the statistical inventory.

Refs. 1 through 5 give a rather sound base on which to expand available knowledge concerning order statistics. Harter (Refs. 1, 2, and 3) apparently had planned a series of volumes on the general applicability of sample order statistics to various Department of Defense (DOD) problems, but with the appearance of Harter's Ref. 3, it becomes clear that some adjustments and changes were necessary in view of the passage of time and the very wide scope of research into order statistic theory by many different investigators. One of the motivating forces behind the publication of Ref. 1 was the need to bring together and summarize much useful information on multiple comparison tests, for example, to establish superiority of one treatment over another in the analysis of variance (ANOVA). Hence Ref. 1, which was aimed at treating order statistics and their use in testing and estimation, discusses and gives rather complete tables of the range and "Studentized" range in random samples from a normal population. The Studentized range is defined as the quantity used for outlier tests in Chapter 3

$$t = w/s \tag{7-2}$$

where

w = sample range as in Eq. 7-1

s = standard deviation of the sample, usually based on $(n - 1)$ degrees of freedom (df).

Harter's Ref. 2 carries on with his original plans and discusses estimates of population parameters based on order statistics from various types of populations, including the normal, exponential, Weibull, gamma, and extreme value distributions. A very useful introduction with many important references is included in Ref. 2.

With the publication of Harter's Ref. 3 in 1977, it became entirely obvious that the field of order statistics had grown so extensively and rapidly that Harter's original plan to cover the many important and useful topics on order statistics had to be abandoned in favor of a chronological and annotated bibliography. Vol. 1 (Ref. 3) of the new series covers topics of interest by various authors for the the pre-1950 time period. Presumably, this type of chronological and annotated bibliography will continue at least into the immediate future.

Sarhan and Greenberg's *Contributions to Order Statistics* (Ref. 4), which was first published in 1962, served more or less as the accepted standard on state of the art coverage of order statistics for many years, and along with David's book (Ref. 5) any serious reader has available in these two volumes much of the theory and many of the topics he will have occasion to use. As David (Ref. 5) points out in the Preface, his book is not intended to replace the Sarhan-Greenberg book (Ref. 4) because the tables of the latter book, and indeed much of the theory, will continue to remain very useful for many years to come. In fact, Ref. 4 is often used as a valuable handbook for reference purposes, and many of the tables from a large number of sources are sufficiently complete. David's book should be considered as an update of theoretical contributions to order statistics and also perhaps as a useful textbook that cites many, many references on order statistic topics up through about 1969.

For this handbook we consider our goal to be that of highlighting some of the material available in Refs. 1-5 and, more importantly, to supplement it—especially to record certain topics in order statistic theory that may often be of value in Army applications. To this end, we will give a brief account of some of the key distributions, the estimation of parameters from truncated or censored samples, some appropriate applications on confidence bounds, and the relation of order statistic theory to general statistical theory.

7-2 THE DISTRIBUTION OF THE LARGEST AND SMALLEST SAMPLE VALUES, THE DISTRIBUTION OF THE RANGE, AND THE rth ORDER STATISTIC

The probability distributions of the largest observation and the smallest observation in samples of size n from any general statistical population with probability density function (pdf) of $f(x)$ and cumulative distribution function (cdf) of $F(x)$ are easily obtained. In fact, for the largest sample value we merely are determining the chance that all the sample observations do not exceed the largest sample value x_n, which is clearly given by the expression

$$Pr[x_n \leq x] = F_n(x) = Pr[\text{all } x_i \leq x] = [F(x)]^n. \tag{7-3}$$

In a very like manner, the cdf of the smallest value x_1 is simply

$$F_1(x) = 1 - Pr[\text{all } x_i > x] = 1 - [1 - F(x)]^n. \tag{7-4}$$

Differentiation of Eqs. 7-3 and 7-4 gives the appropriate pdf's if desired. Also it is readily seen that, given any value of x, the cumulative probability of either extreme sample value can be obtained for a specified $F(x)$ and that the inverse problem to find x for a given level of probability can be easily calculated. In Ref. 6 Tippett first gave distributional properties of the extreme individuals in samples of n from a normal population, and he also tabulated moment properties and the probability distribution of the range $w = x_n - x_1$, which we derive next.

For the distribution of the sample range w, one may see that no matter what the value of a random variable x, the chance that just one of the x_i falls into the interval $(x, x + dx)$ and all of the remaining $(n - 1)$ sample values x_i fall into the interval $(x, x + w)$ is the quantity given by the expression

$$nf(x)dx[F(x + w) - F(x)]^{n-1}. \tag{7-5}$$

To find the cumulative probability distribution of the sample range w, one integrates x over its range of values. Thus the cdf $F(w)$ of the sample range w is

$$F(w) = n\int_{-\infty}^{\infty} f(x)\,[F(x + w) - F(x)]^{n-1}dx. \tag{7-6}$$

Moment constants of the range and the probability integral of the range are given in Refs. 1, 6, and 7 as are tables of percentage points.

How to obtain the probability distributions of the least sample value, the greatest sample observation, and the range for any general population with cdf of $F(x)$ having been indicated, it is also a straightforward matter to derive the cdf of the rth sample order statistic. Thus if we set $i = r$ to designate the rth ordered sample value, the distribution of x_r may be obtained by finding the chance that at least r of the observed x_i's are less than or equal to a value x, and this is

$$\begin{aligned} F_r(x) &= \sum_{i=r}^{n} \binom{n}{i} [F(x)]^i[1 - F(x)]^{n-i} \\ &= I_{F(x)}(r, n - r + 1) \\ &= Pr\{F(2n - 2r + 2, 2r) > r[1 - F(x)]/[(n - r + 1)\,F(x)]\}^* \end{aligned} \tag{7-7}$$

where $I_y(u,v)$ is Karl Pearson's incomplete beta function (Ref. 7), and the quantity $F(u,v)$ is the Fisher-Snedecor F statistic with u and v df, respectively. Therefore, one easily recognizes the cdf of the rth sample order statistic as a sum of binomial terms, i.e., the upper $(n - r)$ or last terms.

In summary, therefore, we find that fairly elementary probability distributions will characterize the chance distributions of either the least observation, the greatest one, the sample range, or the rth sample order statistic for any general population $F(x)$. Of course, in particular applications one would select for $F(x)$ the normal distribution, the exponential distribution, the gamma distribution, or the Weibull distribution, etc., depending on which law best fits the data at hand. In reliability and life testing, for example, the exponential or Weibull models most likely would be the proper ones to apply.

For many distributions it becomes a rather easy matter to find the value of x in Eq. 7-7 that will determine any percentage point or quantile of the distribution of the rth order statistic in a sample of size n. For example, if we let $x(\alpha)$ be the value of x that corresponds with the α probability level of the distribution of the rth order statistic, we will illustrate by Example 7-1 just how any quantile of the exponential distribution with mean failure time of θ may be found.

Example 7-1:

Given that $F(x) = 1 - \exp(-x/\theta)$ or that $F(x)$ is the cumulative exponential distribution, as in problems for life testing or reliability, find the 90% probability level for x_r, the rth order statistic in a sample of size n.

For any general cdf Guenther (Ref. 8) has suggested a rather simple procedure for determining the quantity $x(\alpha)$ desired. First, it is noted that the quantity within braces on the right-hand side (RHS) of the last line of Eq. 7-7 means that for equality

$$F_r[x(\alpha)] = \frac{r}{r + (n - r + 1)F_{1-\alpha}(2n - 2r + 2, 2r)} = q \tag{7-8}$$

*The reader should note in Eq. 7-7 that $F(x)$ is a cdf, whereas $F(u,v)$ is the "F" statistic.

from which—for any given sample size n and r—we can for any stated probability level α substitute the value of "F" for $(2n-2r+2)$ and $2r$ df and then can determine the numerical value of the quantity q. With the value of q so determined, it is a straightforward matter to find the percentage point $x(\alpha)$ for the exponential distribution for we have

$$1-\exp(-x/\theta)=q \tag{7-9}$$

so that solving for x we have

$$x=x(\alpha)=-\theta\ln(1-q). \tag{7-10}$$

To complete the solution, we have $\alpha=0.90$, and for any rth order statistic in a sample of size n, one uses r, n, and the value of $F_{0.10}(2n-2r+2,2r)$ in Eq. 7-8 to find q. Finally, the value of $x(0.90)$ is the negative of the mean life θ multiplied by the natural logarithm of $(1-q)$ as in Eq. 7-10. For other distributions, such as the normal distribution, for example, interpolation or cut-and-try methods may be used as necessary.

For interested readers Guenther (Ref. 8) gives general solutions in terms of the quantity q for the standard logistic distribution, the lognormal distribution, the double exponential distribution, the Pareto distribution, and the "standard" (one-parameter) Weibull model. He also indicates solutions by trial for the normal, the lognormal, the gamma, and the Cauchy distributions. For discrete distributions Guenther (Ref. 8) discusses the binomial and the Poisson distributions.

In addition to the distributional properties of order statistics for any general model or cdf, the moment properties of order statistics are also of considerable interest in applications. Thus the mean, the variance or standard deviation, and often the skewness coefficient and the kurtosis coefficient represent parameters of importance. These moment properties of the order statistics depend numerically on the particular population sampled so that the construction of tables of such values is necessary except in the simplest analytical cases. In fact, this is primarily the reason for the publication of many of the extensive tables in Refs. 1 and 2 of Harter and also for many of the tabulations given in Sarhan and Greenberg (Ref. 4).

In addition to the lower moment properties of the sample order statistics, we should also mention the so-called "quasi-ranges", which involve the inner ordered observations of the sample and hence should not be sensitive to outliers. Therefore, we will discuss the quasi-ranges very briefly and then proceed to some limited account of moments of order statistics.

7-3 THE QUASI-RANGES

Often it could be very desirable to avoid using the extreme values in samples for estimation purposes since the sample range, for example, includes both the largest and the smallest sample values and could be sensitive to the existence of outliers. It is for this and other reasons that some investigators have investigated the properties of quasi-ranges. The rth quasi-range is defined as the quantity

$$w_r=x_{n-r}-x_{r+1} \tag{7-11}$$

or, that is, the $(n-r)$th order statistic minus the $(r+1)$st sample order statistic. If r is set equal to zero, then w_0 of Eq. 7-11 becomes the ordinary sample range defined in Eq. 7-1. As is the case for the range of the complete sample, the rth quasi-range may be used with proper divisor or multiplication factor to give a quick estimate of the population sigma or standard deviation. There is the question of just which quasi-range—i.e., $r=0$, 1, 2, . . ., etc.—should be used for estimation purposes. This particular problem has been studied by Cadwell (Ref. 9), who discovered that for samples of size up through $n=17$, $w_0=x_n-x_1$ should be used, beyond which sample size w_1 becomes optimum through a sample size of $n=31$, where w_2 becomes better, etc.; these results are for normal populations. Table A5 of Harter (Ref. 2) gives the most efficient point estimators of the normal population sigma or standard deviation for samples of size $n=2(1)100$.

Harter's Table A1 of Ref. 2 gives the means or expected values of quasi-ranges numerically for sample sizes of $n=2(1)100$ and values of r less than or equal to the sample size n. Table A2 gives the variances of the quasi-ranges for the same conditions on the sample size n and order r.

The cumulative probability integral of the rth quasi-range for selected sample sizes up through $n = 100$, and the percentage points of these quasi-ranges, for normal samples are given in Harter's Tables A6 and A7, respectively, of Ref. 2. Also his introductory discussions give necessary details concerning the tables and methods of computation.

We refer to Harter's tables primarily here for they are the most extensive and most readily available from the Government Printing Office (GPO).

The discussion here of the use of quasi-ranges in samples brings forth the idea that there may be other methods of treating sample observations to obtain efficient estimates of the population standard deviation for normal samples. In fact, instead of dealing with quasi-ranges of the large samples, one might consider dividing the entire sample into a number of subgroups and then using the average range of the subgroups to obtain a more precise or efficient estimate of the normal population standard deviation. The size of the subgroups becomes of importance in the division of large samples for such purposes, and the problem has been studied by Grubbs and Weaver (Ref. 10). They found that subgroups of size about eight were the most efficient ones to use, so large samples are divided accordingly with an occasional size of seven or nine permitted. As an example, Ref. 10 discusses the estimation of a normal population sigma for a sample of size 30. In this case, one uses two subgroups of size seven and two of size eight. See Ref. 10 for further details.*

A very important point we should bring out in connection with the use of order statistics is that the range, the average range, the individual order statistic, and the quasi-ranges all have to be multiplied by appropriate numerical factors to make them unbiased estimates of the mean, standard deviation, and other parameters of populations. Moreover, therefore, it is seen that it becomes very natural to use linear functions of the order statistics to estimate any parameter of the population sampled. Thus it should be expected that linear estimation principles tie in directly with the use of sample order statistics. In addition, it is observed also that once a weighted linear function of the sample order statistics is used, the matter of finding its variance becomes rather straightforward since such variances will depend on the coefficients or weighting factors, the variances of the order statistics, and the covariances of the ordered sample values. Hence we see the importance of linear estimation principles. Since the expected or mean values and the higher moments of the sample order statistics are needed in connection with linear estimation methods, and indeed are easily found, we will discuss this topic next.

7-4 EXPECTED VALUES AND MOMENTS OF SAMPLE ORDER STATISTICS

The means or expected values and all of the moments of the order statistics are rather easily found since the pdf of the rth order statistic may be determined from the RHS of the first line of Eq. 7-7 by differentiation. Thus we see that

$$f_r(x) = n\binom{n-1}{r-1}[F(x)]^{r-1}[1 - F(x)]^{n-r}dF(x)/dx. \tag{7-12}$$

Furthermore, the kth moment about the origin is found from the expression

$$E(x_r^k) = \int_{-\infty}^{\infty} x_r^k f_r(x)dx \tag{7-13}$$

where we have used Eq. 7-12. Therefore, with a given sample size n, the order r of the sample statistic desired, and the functional form $F(x)$ of the distribution of interest, the moments about the origin of the rth order statistic may be calculated from Eq. 7-13, especially with the aid of a computer. The central moments then may be calculated with the usual conversion equations given in standard statistical textbooks. For example, for $k = 1$ in Eq. 7-13, one determines the population mean or the expected value of the rth order statistic. For $k = 2$ the second moment about the origin is determined, and if the square of the mean or expected value is subtracted from this second moment about the origin, the result gives the variance of the rth order statistic. The third and fourth central moments are used to find the skewness and kurtosis, respectively.

*Quasi-ranges are often more efficient than the mean or average range—see Harter (Ref. 2).

7-5 LINEAR ESTIMATION OF POPULATION PARAMETERS OR MOMENTS

As we have indicated in par. 7-3, it becomes highly desirable to use weighted linear functions of the sample order statistics to estimate the population mean, variance, or standard deviation, etc., or higher moments. The use of linear functions to estimate parameters avoids complications of other types of estimation techniques, such as maximum likelihood (ML) estimation, for example. Thus tables of coefficients by which to multiply each of the order statistics, or some of them, and to sum the results would lead to very acceptable estimators of parameters provided they are efficient enough. As it turns out, linear estimation in connection with order statistics leads to very efficient estimators for many important populations of interest in practice—such as the normal population, the lognormal population, the exponential population, the gamma population, and the Weibull type of model. The linear estimators are very efficient provided the distribution function has a form such that it may be expressed in terms of a linear function of the population mean and standard deviation.

A number of authors have studied linear estimation using the sample order statistics, including especially E. H. Lloyd (Ref. 11) whose generalized least squares theorem is also given in Chapter 3 of Sarhan and Greenberg's book (Ref. 4). Primary emphasis is on the estimation of the population mean and the standard deviation or variance. Also the coefficients are usually determined so that the linear estimators are unbiased and have minimum variance, or they could be determined so that the minimum mean square error (MSE) is guaranteed, etc. Efficient estimators are often referred to as "BLUE" or "best linear unbiased estimators", and these are the primary ones that have been determined and tabulated for various populations of practical importance.

It is not within the scope of this handbook to give a very extensive account of the theory or other details of the best linear estimation techniques; interested readers may consult Sarhan and Greenberg's book (Ref. 4), David's book (Ref. 5), the various references of this chapter, and David's very extensive coverage of references on order statistics on pp. 235-66 of Ref. 5. Here we will indicate the general nature of the equations for the mean and standard deviation and will follow this with a discussion of the necessary tables and some examples.

The population mean μ is estimated by a linear form of the type

$$\text{est}\,\mu = \Sigma a_r x_r \tag{7-14}$$

where

a_r = constant or coefficient related to rth sample order statistic

and where the sum may be taken over the whole sample $r = 1, 2, \ldots, n$, or only over (the inner) part of the sample order statistics. In a like manner, the estimator of the population sigma is found by using a similar sum involving different coefficients or

$$\text{est}\,\sigma = \Sigma b_r x_r \tag{7-15}$$

where

b_r = constant or coefficient related to rth sample order statistic

for which some of the end points may be truncated or censored. Our primary interest will be in the BLUE estimators.

7-6 DISCUSSION OF TABLES AND SOME EXAMPLES

To use sample order statistics, it is absolutely necessary to have tables of the coefficients available. For all of the applications analysts are likely to face in practice, the tables of coefficients amount to literally hundreds of pages. Therefore, it cannot be expected that any extensive coverage of the tables can be displayed in this chapter. Nevertheless, we can give an example and make references to and discuss some of the types of tables that are available.

As mentioned before, Harter's Ref. 1 gives a very extensive set of tables for the sample range and its properties. Ref. 1 covers the probability integral of the range, the percentage points of the range, and the

moments of the range and includes the percentage points of the ratio of two ranges and related tables. The ratio of two ranges may, of course, be used just like the Fisher-Snedecor F ratio to judge whether the variances of two normal populations are equal.

For the Studentized range, Ref. 1 gives both the probability integral and the percentage points as well as critical values of Duncan's (Ref. 12) multiple range tests for judging contrasts in an ANOVA. Moreover, instructions and examples are given in Harter's introductory sections of Ref. 1. On p. 30 of the "Introduction" to Chapter 2 of Ref. 1, the determination of sample sizes for the multiple range tests is discussed. This brief discussion may give the reader some idea of the value of Harter's Ref. 1.

Example 7-2:

A velocity dispersion test was conducted to determine whether a new technique to apply rotating bands to artillery projectiles was superior to the standard method. Fifteen projectiles, with rotating bands applied with the new technique, were fired along with 15 reference projectiles, and the velocities were measured. The new technique gave a range in velocity dispersion of 9 ft/s, and the standard projectiles had a range in velocity dispersion of 15 ft/s. Does the new technique give a smaller standard deviation in velocity? Assume normal populations.

The ratio of the two ranges in velocity dispersion, which we will call F', is given by

$$F' = 15/9 = 1.667.$$

Referring to Harter's Table A4 of Ref. 1 for the percentage points of the ratio of two ranges with sample sizes $n_1 = n_2 = 15$, one finds on p. 227 of Ref. 1 that the 95% level of F' is 1.673. Hence the result is beginning to appear significant. It might be advisable, however, if a costly decision is being made, to fire a larger number of rounds for final judgment.

Example 7-3:

Use the observed data of Example 7-2 to estimate population sigmas, assuming normal parents.

From Harter's Table A8 of Ref. 1, p. 376, one finds that the expected value $E(w)$ of a range for a sample of size 15 is

$$E(w) = 3.4718268899\sigma.$$

Hence the estimated standard deviations of the populations for both projectiles are $9/3.472 = 2.6$ and $15/3.472 = 4.3$ ft/s, respectively.

Harter's Volume 2 on order statistics and their use in testing and estimation (Ref. 2) contains many useful tables for applications to a variety of Army statistical problems. Both point and interval estimation of the normal population standard deviation using the quasi-ranges are covered, along with the probability integral of quasi-ranges, percentage points, and efficiencies of the best choices of quasi-ranges. The range of samples chosen at random from a rectangular population is covered, including both point and interval estimation. Also the percentage points of the range for a rectangular parent are given in Harter's Table B3, p. 415 , of Ref. 2. These percentage points are for sample sizes 1(1)20(2)40(10)100. Coefficients of the range for the same sample sizes for exact lower confidence bounds on the rectangular population standard deviation also are presented in Table B4 of Harter's Ref. 2.

Expected values of the order statistics for samples of size n drawn from a normal population, an exponential population, a Weibull parent, and a gamma universe are given in Appendix C of Ref. 2. It is believed that such tables will be very useful. Moments of the sample order statistics are tabulated for the exponential, Weibull, and gamma populations in Table C5 of Harter's Ref. 2 for certain values of the shape parameter for the cases of Weibull or gamma populations. Since this is only a one-page table, we are including these moment constants here as Table 7-1 because such properties will have interest on occasion. In contrast with the standard normal population, we recall for this case that the mean is zero, the variance is one, the skewness is zero, and the kurtosis is three. Hence we note that the exponential, Weibull, and gamma populations can be decidedly skewed and peaked.

TABLE 7-1

MOMENTS OF EXPONENTIAL, WEIBULL, AND GAMMA POPULATIONS

Population	Shape Parameter	Mean	Variance	Skewness	Kurtosis
Exponential		1.00000000	1.00000000	2.00000000	9.00000000
Weibull	0.5	2.00000000	20.00000000	6.61876121	87.72000000
Weibull	1.0	1.00000000	1.00000000	2.00000000	9.00000000
Weibull	1.5	0.90274529	0.37569028	1.07198657	4.39040356
Weibull	2.0	0.88622693	0.21460184	0.63111066	3.24508930
Weibull	2.5	0.88726382	0.14414669	0.35863184	2.85678309
Weibull	3.0	0.89297951	0.10533288	0.16810284	2.72946363
Weibull	3.5	0.89974718	0.08107275	0.02510816	2.71273189
Weibull	4.0	0.90640248	0.06466148	−0.08723697	2.74782953
Weibull	5.0	0.91816874	0.04422998	−0.25410959	2.88029006
Weibull	6.0	0.92771933	0.03231635	−0.37326156	3.03545528
Weibull	7.0	0.93543756	0.02470374	−0.46318962	3.18718296
Weibull	8.0	0.94174270	0.01952316	−0.53372638	3.32767551
Gamma	0.5	0.50000000	0.50000000	2.82842712	15.00000000
Gamma	1.0	1.00000000	1.00000000	2.00000000	9.00000000
Gamma	1.5	1.50000000	1.50000000	1.63299316	7.00000000
Gamma	2.0	2.00000000	2.00000000	1.41421356	6.00000000
Gamma	2.5	2.50000000	2.50000000	1.26491106	5.40000000
Gamma	3.0	3.00000000	3.00000000	1.15470054	5.00000000
Gamma	3.5	3.50000000	3.50000000	1.06904497	4.71428571
Gamma	4.0	4.00000000	4.00000000	1.00000000	4.50000000

For this table the location parameters are taken as zero. Also as is applicable, the scale and/or shape parameters are taken to be unity. Thus the cdf's of the exponential, Weibull, and gamma models are

Exponential: $F(x) = 1 - \exp(-x/\theta)$, $\theta = 1$
Weibull: $F(x) = 1 - \exp[-(x/\theta)^{\beta}]$, $\theta = 1$, β varies
Gamma: $F(x) = \int_0^x x^{\beta} \exp(-x/\theta)\, dx/(\beta!\theta^{\beta+1})$, $\theta = 1$, β varies.

Due to the large number of pages involved, we cannot list the expected values of the sample order statistics for all populations or sample sizes of practical interest. Nevertheless, in Table 7-2 we give the expected values of the sample order statistics for samples of size 2(1)20 for the standardized normal parent. The tabular values are taken from Teichroew's paper (Ref. 13). The reader should note in particular that only the lower expected values of the order statistics are listed; accordingly, all table entries should be preceded by a negative sign. The values of i for order statistics above the median would have positive signs, as seen by the example given at the bottom of Table 7-2. The entries in Table 7-2 are for a normal population with zero mean and standard deviation of unity. Therefore, if one is sampling a normal population with mean μ and standard deviation σ, the values in Table 7-2 must be multiplied by σ, or an estimate of σ, when making inferences about the sampled population.

Harter's Table C1, p. 425, Ref. 2, of the expected values of normal order statistics is very extensive; it extends through a sample of size 400 (with some missing intermediate values). The tabular entries of Table C1,

DARCOM-P 706-103

TABLE 7-2

EXPECTED VALUES OF ORDER STATISTICS FROM $N(0,1)$ (Ref. 13)

n	r	$E(x_r,n)$	n	r	$E(x_r,n)$	n	r	$E(x_r,n)$
2	1	0.56418 95835	12	5	0.31224 88787	17	5	0.61945 76511
3	1	0.84628 43753	12	6	0.10258 96798	17	6	0.45133 34467
4	1	1.02937 53730	13	1	1.66799 01770	17	7	0.29518 64872
4	2	0.29701 13823	13	2	1.16407 71937	17	8	0.14598 74231
5	1	1.16296 44736	13	3	0.84983 46324	18	1	1.82003 18790
5	2	0.49501 89705	13	4	0.60285 00882	18	2	1.35041 37134
6	1	1.26720 63606	13	5	0.38832 71210	18	3	1.06572 81829
6	2	0.64175 50388	13	6	0.19052 36911	18	4	0.84812 50190
6	3	0.20154 68338	14	1	1.70338 15541	18	5	0.66479 46127
7	1	1.35217 83756	14	2	1.20790 22754	18	6	0.50158 15510
7	2	0.75737 42706	14	3	0.90112 67039	18	7	0.35083 72382
7	3	0.35270 69592	14	4	0.66176 37035	18	8	0.20773 53071
8	1	1.42360 03060	14	5	0.45556 60500	18	9	0.06880 25682
8	2	0.85222 48625	14	6	0.26729 70489	19	1	1.84448 15116
8	3	0.47282 24949	14	7	0.08815 92141	19	2	1.37993 84915
8	4	0.15251 43995	15	1	1.73591 34449	19	3	1.09945 30994
9	1	1.48501 31622	15	2	1.24793 50823	19	4	0.88586 19615
9	2	0.93229 74567	15	3	0.94768 90303	19	5	0.70661 14847
9	3	0.57197 07829	15	4	0.71487 73983	19	6	0.54770 73710
9	4	0.27452 59191	15	5	0.51570 10430	19	7	0.40164 22742
10	1	1.53875 27308	15	6	0.33529 60639	19	8	0.26374 28909
10	2	1.00135 70446	15	7	0.16529 85263	19	9	0.13072 48795
10	3	0.65605 91057	16	1	1.76599 13931	20	1	1.86747 50598
10	4	0.37576 46970	16	2	1.28474 42232	20	2	1.40760 40959
10	5	0.12266 77523	16	3	0.99027 10960	20	3	1.13094 80522
11	1	1.58643 63519	16	4	0.76316 67458	20	4	0.92098 17004
11	2	1.06191 65201	16	5	0.57000 93557	20	5	0.74538 30058
11	3	0.72883 94047	16	6	0.39622 27551	20	6	0.59029 69215
11	4	0.46197 83072	16	7	0.23375 15785	20	7	0.44833 17532
11	5	0.22489 08792	16	8	0.07728 74593	20	8	0.31493 32416
12	1	1.62922 76399	17	1	1.79394 19809	20	9	0.18695 73647
12	2	1.11573 21843	17	2	1.31878 19878	20	10	0.06199 62865
12	3	0.79283 81991	17	3	1.02946 09889			
12	4	0.53684 30214	17	4	0.80738 49287			

(The i in Teichroew's table has been replaced by r.)

For the values of r in the table, all entries should be preceded by a negative sign since the r's are for sample order statistics below the sample median.

Example:

$E(x_3,10) = -0.65606$

but

$E(x_8,10) = +0.65606.$

Reprinted with permission. Copyright © by Institute of Mathematical Statistics.

Ref. 2, are given to five decimal places and hence should cover practically all needs. These are based on Eq. 7-13 with $i = r$, $k = 1$, and $F(x) = \int_{-\infty}^{x} \exp(-x^2/2)\, dx/\sqrt{2\pi}$.

Blom (Ref. 14) points out that a rather good approximation to the expected value of the rth normal sample order statistic may be determined from the relation

$$E(x_r) = F^{-1}[(r - 3/8)/(n + 1/4)] \tag{7-16}$$

where

$F(x) = \int_{-\infty}^{x} \exp(-x^2/2)dx/\sqrt{2\pi}$.

Expected values of the sample order statistics for the exponential, the Weibull, and the gamma distributions are tabulated by Harter in his Tables C2, C3, and C4, respectively, Ref. 2. The sample sizes covered for the exponential population are up through $n = 120$; for the Weibull and gamma parents the sample sizes are through $n = 40$.

Appendix D of Harter's Volume 2 is devoted to tables for one- and two-order statistic estimators for exponential populations and include:

1. Table D1. Most Efficient Unbiased Point Estimators for σ, Based on One- and Two-Order Statistics of a Sample from a One-Parameter Exponential Population
2. Table D2. Unbiased Point Estimators for σ, Based on One-Order Statistic of a Censored Sample from a One-Parameter Population
3. Table D3. Most Efficient Unbiased Point Estimators for Two Parameters, Based on Two-Order Statistics of a Two-Parameter Exponential Population
4. Table D4. Most Effective (Efficient) Interval Estimators for σ, Based on One-Order Statistic of a Sample from a One-Parameter Exponential Population.

Appendix E of Ref. 2 gives tables of conditional ML estimators from singly censored samples. The coverage in particular includes:

1. Table E1. Weibull Population—Unbiasing Factors and Variances of Unbiased Estimators
2. Table E2. Type I Extreme-Value Population*—Biases and Variances of Unbiased Estimators
3. Table E3. Type II Extreme-Value Population—Unbiasing Factor, Variance, and Efficiency.

Finally, Appendix F of Harter (Ref. 2) gives tables related to the asymptotic variances and covariances of ML estimators from doubly censored samples, and Appendix G covers some tables of results of Monte Carlo studies of ML estimators from doubly censored samples.

Clearly, and in summary, the Army analyst should find Refs. 1 and 2 by Harter to be necessary aids in the analysis of sample order statistics and in related applications.

Harter's tables in Refs. 1 and 2, although very extensive in nature, do not encompass all such requirements. Rather, there are many tables in Sarhan and Greenberg's book (Ref. 4) and elsewhere that will be required, depending on the particular application. For example, suppose that one acquires singly or doubly truncated samples from a normal, exponential, Weibull, or gamma population and desires to estimate the mean and standard deviation using the BLUE. He will need the coefficients for the BLUE for the particular population he is sampling, as discussed initially in par. 7-5. With regard to this general type of problem, we give in Table 7-3 the coefficients for the BLUE for a normal population, which often may be used in applications. These coefficients are given for sample sizes up through $n = 10$ and for singly and doubly truncated samples. The coefficients in Table 7-3—which are used with observed sample order statistics to give the minimum variance, unbiased linear estimators of the the normal population mean and sigma—are taken from Table II of Sarhan and Greenberg's paper (Ref. 15). For values of the sample size n through 20, see Table 10C.1 of Sarhan and Greenberg's book (Ref. 4). In Table 7-3 r_1 is the number of smallest ordered sample values censored, and r_2 is the number of largest sample observations censored, in the total sample size n. (In Table 7-3, there are 10 columns for the x_i.) The upper values listed in Table 7-3 are for estimation of the normal population mean; the lower entries are for coefficients to estimate the normal population sigma. Example 7-4 follows.

Example 7-4:

Ten experimental projectiles were fired at a 6-ft by 6-ft vertical target, and the impact points, or holes, as measured from the left-hand edge were at 11, 26, 41, 56, and 70 in. The gunner noted that one projectile missed the target on the left, and four rounds hit the ground on the right side of the target. Nevertheless, determine estimates of the mean horizontal point of impact, or center of impact (C of I), and the round-to-round standard deviation by assuming a normal distribution of impacts.

*For the extreme-value model, $F(x) = 1 - \exp[-\exp(-x/\beta)]$.

Since five rounds missed the target, one on the left and four on the right, we have

$$n = 10, r_1 = 1, \text{ and } r_2 = 4.$$

Referring to Table 7-3 for these conditions, we note the coefficients for the BLUE of the mean and standard deviation, so one may calculate immediately

$$\hat{\mu} = -0.0043(11) + 0.0665(26) + 0.0938(41) + 0.1179(56) + 0.7261(70) = 62.96 \text{ in.}$$

$$\hat{\sigma} = -0.7359(11) - 0.1719(26) - 0.0797(41) + 0.0031(56) + 0.9844(70) = 53.25 \text{ in.}$$

Note that for estimating the population mean, the largest distance to the right-hand shot on the target carries 73% of the weight; consequently, the mean is estimated to be somewhat near the RHS of the target. For estimation of the normal population sigma, the second sample order statistic uses a relative weight of 0.74 versus the sixth order statistic, which has a relative weight of 0.98; the ratio is 0.98/0.74 = 1.32. (The sum of the weights for sigma add to unity.) In any event we see that the normal population sigma is estimated to be quite large, or about 53/(6 x 12) = 74% of the target width because so many rounds missed the target. The advantage of the order statistics is, of course, that the population parameters can still be estimated in an unbiased manner even though half the data are missing!

As pointed out by Sarhan and Greenberg in Ref. 15, coefficients may be determined for values of r_1 and r_2 not given in their tables:

"If the coefficients of an estimate are sought for a value of r_1 not given in the table, these can be obtained by interchanging the values of r_1 and r_2 and rearranging the observations in descending order. In such an event, the coefficients for the best linear systematic statistic of the mean will be identical with those given in the table, whereas those for the standard deviation will be numerically the same but with opposite sign.".

With reference to coefficients of the BLUE for the exponential, Weibull, and other populations, the reader should consult Refs. 4 and 5.

7-7 SOME RELATIONS AND USES OF ORDER STATISTICS WITH RESPECT TO ALLIED STATISTICAL PROBLEMS

7-7.1 SOME PARTICULAR USES OF ORDER STATISTICS

David (Ref. 16) discusses some particular uses of the sample order statistics in connection with system reliability, the problem of "data compression", some selection procedures, and double sampling. We will indicate some of these applications.

Suppose we have a parallel system of n components, which are alike and for which each component follows the same time-to-fail law with any general cumulative distribution function $F(x)$. Thus if $x_{(i)}$ represents the time-to-fail of the ith component of the parallel system, the largest observation, or failure time, x_n also represents the failure time of the entire parallel system. Thus the cdf of the system will be given by Eq. 7-3 or $[F(x)]^n$, or the distribution of the largest component lifetime.*

In a like manner, the least sample value may be used to describe the lifetime of a series system of similar components for here the chance that all component lifetimes exceed any given failure time x is $[1 - F(x)]^n$ as contrasted to Eq. 7-4**. Thus we are able to deduce the probability distributions of series and parallel system lifetimes.

Furthermore, as pointed out by David in Ref. 16—even though the components may have different failure-time distributions, which we will represent here as $P_i(x)$ for the ith component—the lifetime probability distribution of the parallel system will be given by

$$Pr[x_n \leq x] = \prod_{i=1}^{n} P_i(x)]. \tag{7-17}$$

*The "reliability" $R(x)$ of the parallel system is $R(x) = 1 - [F(x)]^n$.

**The quantity (Eq. 7-4) is the reliability of the series system.

TABLE 7-3

THE COEFFICIENTS OF THE MOST EFFICIENT LINEAR SYSTEMATIC STATISTICS OF THE MEAN AND STANDARD DEVIATION IN CENSORED SAMPLES OF SIZES ≤10 FROM A NORMAL POPULATION (Ref. 15)

n	r_1	r_2	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
2	0	0	0.50000000	0.50000000								
			—0.88622693	0.88622693								
3	0	0	0.33333333	0.33333333	0.33333333							
			—0.59081795	0.00000000	0.59081795							
	0	1	0.00000000	1.00000000								
			—1.18163590	1.18163590								
4	0	0	0.25000000	0.25000000	0.25000000	0.25000000						
			—0.45394040	—0.11018073	0.11018073	0.45394040						
	0	1	0.11606577	0.24083805	0.64309018							
			—0.69713303	—0.12681665	0.82394968							
	0	2	—0.40555159	1.40555159								
			—1.36544125	1.36544125								
	1	1		0.50000000	0.50000000							
				—1.68343717	1.68343717							
5	0	0	0.20000000	0.20000000	0.20000000	0.20000000	0.20000000					
			—0.37238157	—0.13521392	0.00000000	—0.13521392	0.37238157					
	0	1	0.12515679	0.18304590	0.21471643	0.47708089						
			—0.51173274	—0.16678091	0.02740065	0.65111300						
	0	2	—0.06377484	0.14982836	0.91394649							
			—0.76958387	—0.21211572	0.98169958							
	0	3	—0.74110683	1.74110683								
			—1.49712813	1.49712813								
	1	1		0.38929103	0.22141794	0.38929103						
				1.01006230	0.00000000	1.01006230						
	1	2		0.00000000	1.00000000							
				—2.02012460	2.02012460							
6	0	0	0.16666667	0.16666667	0.16666667	0.16666667	0.16666667	0.16666667				
			—0.31752484	—0.13855961	—0.04321165	0.04321165	0.13855961	0.31752484				
	0	1	0.11828773	0.15097353	0.16803141	0.18280232	0.37990501					
			—0.40969394	—0.16845737	—0.04061162	0.07395248	0.54481044					
	0	2	0.01848367	0.12260668	0.17614875	0.68276091						
			—0.55281996	—0.20913743	—0.02897078	0.79092817						

(cont'd on next page)

TABLE 7-3 (cont'd)

n	r_1	r_2	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
6	0	3	—0.21591800	0.06485167	1.15106633							
			—0.82435700	—0.27604235	1.10039935							
	0	4	—1.02606712	2.02606712								
			—1.59884545	1.59884545								
	1	1		0.31977001	0.18022999	0.18022999	0.31977001					
				—0.75309024	—0.08286186	0.08286186	0.75309024					
	1	2		0.15385831	0.17811699	0.66802470						
				—1.14382882	—0.08783818	1.23166700						
	1	3		—0.45784434	1.45784434							
				—2.27165234	2.27165234							
	2	2			0.50000000	0.50000000						
					—2.48081297	2.48081297						
7	0	0	0.14285714	0.14285714	0.14285714	0.14285714	0.14285714	0.14285714	0.14285714			
			—0.27781036	—0.13509780	—0.06246312	0.00000000	0.06246312	0.13509780	0.27781036			
	0	1	0.10882014	0.12954538	0.13997050	0.14873929	0.15705206	0.31587262				
			—0.34400143	—0.16098444	—0.06807697	0.01143886	0.09006788	0.47155609				
	0	2	0.04654966	0.10721153	0.13748095	0.16260139	0.54615647					
			—0.43696302	—0.19432593	—0.07179355	0.03213315	0.67094936					
	0	3	—0.07380239	0.06771901	0.13752310	0.86856027						
			—0.58481466	—0.24284221	—0.07174176	0.89939863						
	0	4	—0.34744564	—0.01345544	1.36090107							
			—0.86817366	—0.32689877	1.19507242							
	0	5	—1.27331716	2.27331716								
			—1.68122579	1.68122579								
	1	1		0.27183155	0.15198954	0.15235782	0.15198954	0.27183155				
				—0.61077842	—0.10607146	0.00000000	0.10607146	0.61077842				
	1	2		0.17480153	0.14322942	0.16338504	0.51858402					
				—0.82879521	—0.12575459	0.02477706	0.92977274					
	1	3		—0.05916442	0.12704486	0.93211956						
				—1.24827431	—0.15477199	1.40304629						
	1	4		—0.87159736	1.87159736							
				—2.47116575	2.47116575							
	2	2			0.41569461	0.16861078	0.41569461					
					—1.41760741	0.00000000	1.41760741					
	2	3			0.00000000	1.00000000						
					—2.83521483	2.83521483						

(cont'd on next page)

TABLE 7-3 (cont'd)

n	r_1	r_2	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
8	0	0	0.12500000	0.12500000	0.12500000	0.12500000	0.12500000	0.12500000	0.12500000	0.12500000		
			—0.24758623	—0.12944776	—0.07130849	—0.02295726	0.02295726	0.07130849	0.12944776	0.24758623		
	0.	1	0.09966946	0.11386201	0.12079601	0.12649213	0.13176484	0.13698715	0.27042840			
			—0.29775817	—0.15150866	—0.07963530	—0.02000181	0.03635632	0.09505131	0.41749631			
	0	2	0.05691876	0.09621316	0.11531993	0.13090112	0.14512418	0.45552284				
			—0.36375811	—0.17875554	—0.08808945	—0.01319507	0.05698091	0.58681726				
	0	3	—0.01672011	0.06765099	0.10840499	0.14131770	0.69934643					
			—0.45862177	—0.21555013	—0.09699747	0.00022386	0.77094550					
	0	4	—0.15491146	0.01760383	0.10013416	1.03717346						
			—0.61096114	—0.27072110	—0.10611506	0.98779730						
	0	5	—0.46316724	—0.08553809	1.54870533							
			—0.90454196	—0.36895282	1.27349478							
	0	6	—1.49153218	2.49153218								
			—1.75016272	1.75016272								
	1	1		0.23666261	0.13147170	0.13186568	0.13186568	0.13147170	0.23666261			
				—0.51837013	—0.11152847	—0.03605506	0.03605506	0.11152847	0.51837013			
	1	2		0.17163950	0.12217411	0.13381130	0.14417854	0.42819655				
				—0.66079246	—0.13189331	—0.03179351	0.06302434	0.76145495				
	1	3		0.04308423	0.10609003	0.14058990	0.71023584					
				—0.88940019	—0.16049536	—0.01973924	1.06963479					
	1	4		—0.25191114	0.07414879	1.17776234						
				—1.33367137	—0.20859974	1.54227111						
	1	5		—1.24622969	2.24622969							
				—2.63572419	2.63572419							
	2	2			0.35694901	0.14305099	0.14305099	0.35694901				
					—1.03574984	—0.06736528	0.06736528	1.03574984				
	2	3			0.17417957	0.14291598	0.68290445					
					—1.56608726	—0.06775706	1.63384432					
	2	4			—0.47614906	1.47614906						
					—3.12199414	3.12199414						
	3	3				0.50000000	0.50000000					
						—3.27837896	3.27837896					
9	0	0	0.11111111	0.11111111	0.11111111	0.11111111	0.11111111	0.11111111	0.11111111	0.11111111	0.11111111	
			—0.22373410	—0.12326850	—0.07509922	—0.03596908	0.00000000	0.03596908	0.07509922	0.12326850	0.22373410	
	0	1	0.09148453	0.10175129	0.10665020	0.11060016	0.11419180	0.11765844	0.12118875	0.23647484		
			—0.26325227	—0.14211324	—0.08407936	—0.03699545	0.00618347	0.04915546	0.09539434	0.37570705		

(cont'd on next page)

TABLE 7-3 (cont'd)

n	r_1	r_2	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
9	0	2	0.06023974	0.08756442	0.10056358	0.11099937	0.12038626	0.12937290	0.39087375			
			—0.31289447	—0.16465459	—0.09375143	—0.03636313	0.01604052	0.06776492	0.52385821			
	0	3	0.01040388	0.06597348	0.09230224	0.11331998	0.13204377	0.58595666				
			—0.37968567	—0.19359128	—0.10482346	—0.03325300	0.03166420	0.67968924				
	0	4	—0.07313367	0.03155502	0.08087391	0.11994592	0.84075879					
			—0.47658635	—0.23351551	—0.11807995	—0.02556713	0.85374893					
	0	5	—0.22717960	—0.02842070	0.06443680	1.19116347						
			—0.63301232	—0.29441786	—0.13477100	1.06220121						
	0	6	—0.56642662	—0.15208218	1.71850879							
			—0.93553052	—0.40469106	1.34022157							
	0	7	—1.68675768	2.68675768								
			—1.80924841	1.80924841								
	1	1		0.20970437	0.11588689	0.11624355	0.11633036	0.11624355	0.11588689	0.20970437		
				—0.45274773	—0.11065463	—0.05323076	0.00000000	0.05323076	0.11065463	0.45274773		
	1	2		0.16260603	0.10736149	0.11483307	0.12137684	0.12749454	0.36632802			
				—0.55443214	—0.12906081	—0.05627598	0.01089528	0.07752140	0.65135225			
	1	3		0.07989097	0.09363515	0.11399182	0.13209087	0.58039119				
				—0.70150427	—0.15346705	—0.05777177	0.02994545	0.88279764				
	1	4		—0.07684209	0.06990479	0.11531222	0.89162509					
				—0.93990137	—0.18956185	—0.05576340	1.18522662					
	1	5		—0.42723635	0.02184604	1.40539031						
				—1.40567629	—0.25344578	1.65912206						
	1	6		—1.58736731	2.58736731							
				—2.77525943	2.77525943							
	2	2			0.31343459	0.12427928	0.12457225	0.12427928	0.31343459			
					—0.83170114	—0.08848435	0.00000000	0.08848435	0.83170114			
	2	3			0.20395900	0.11906520	0.13296497	0.54401083				
					—1.12219549	—0.10231994	0.02227016	1.20224528				
	2	4			—0.05271887	0.10983900	0.94287987					
					—1.68944478	—0.12270952	1.81215430					
	2	5			—0.92294725	1.92294725						
					—3.36196762	3.36196762						
	3	3				0.43149371	0.13701259	0.43149371				
						—1.82132165	0.00000000	1.82132165				
	3	4				0.00000000	1.00000000					
						—3.64264330	3.64264330					

(cont'd on next page)

TABLE 7-3 (cont'd)

n	r_1	r_2	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
10	0	0	0.10000000	0.10000000	0.10000000	0.10000000	0.10000000	0.10000000	0.10000000	0.10000000	0.10000000	0.10000000
			—0.20438349	—0.11719379	—0.07625921	—0.04358325	—0.01421667	0.01421667	0.04358325	0.07625921	0.11719379	0.20438349
	0	1	0.08432557	0.09206394	0.09568489	0.09856130	0.10112450	0.10356728	0.10600993	0.10852533	0.21013724	
			—0.23641943	—0.13341377	—0.08507858	—0.04652372	—0.01191838	0.02150761	0.05586656	0.09368358	0.34229613	
	0	2	0.06045239	0.08044709	0.08978431	0.09718132	0.10374300	0.10994657	0.11605505	0.34239029		
			—0.27530689	—0.15233666	—0.09469015	—0.04877160	—0.00765306	0.03189894	0.07222921	0.47463020		
	0	3	0.02442821	0.06355637	0.08178041	0.09616708	0.10885704	0.12073778	0.50447311			
			—0.32524451	—0.17575100	—0.10578536	—0.05017756	—0.00056384	0.04685799	0.61066429			
	0	4	—0.03158434	0.03829283	0.07072145	0.09620404	0.11851783	0.70784819				
			—0.39304766	—0.20633250	—0.11917222	—0.05013282	0.01113054	0.75755466				
	0	5	—0.12396225	—0.00163103	0.05485447	0.09896448	0.97177434					
			—0.49191253	—0.24905990	—0.13615341	—0.04717854	0.92430437					
	0	6	—0.29230821	—0.07092971	0.03053696	1.33270096						
			—0.65203499	—0.31497342	—0.15928304	1.12629145						
	0	7	—0.65962405	—0.21376659	1.87339063							
			—0.96246073	—0.43568764	1.39814837							
	0	8	—1.86335148	2.86335148								
			—1.86082625	1.86082625								
	1	1		0.18835246	0.10363971	0.10395180	0.10405602	0.10405602	0.10395180	0.10363971	0.18835246	
				—0.40337309	—0.10738113	—0.06163683	—0.02013735	0.02013735	0.06163683	0.10738113	0.40337309	
	1	2		0.15245229	0.09611128	0.10126825	0.10570764	0.10981910	0.11376585	0.32087559		
				—0.48025640	—0.12350392	—0.06738392	—0.01660026	0.03247948	0.08265445	0.57261056		
10	1	3		0.09423044	0.08464767	0.09790957	0.10954691	0.12039960	0.49326580			
				—0.58415474	—0.14396103	—0.07337756	—0.00974899	0.05136065	0.75988167			
	1	4		—0.00426039	0.06648054	0.09383130	0.11787039	0.72607816				
				—0.73588101	—0.17194771	—0.07966018	0.00307342	0.98441548				
	1	5		—0.18661330	0.03511935	0.08919507	1.06229888					
				—0.98311476	—0.21446715	—0.08594598	1.28352789					
	1	6		—0.58772619	—0.02885426	1.61658046						
				—1.46776140	—0.29176358	1.75952498						
	1	7		—1.89997978	2.89997978							
				—2.89604971	2.89604971							
	2	2			0.27977552	0.10994078	0.11028371	0.11028371	0.10994078	0.27977552		
					—0.70208441	—0.09470422	—0.03101584	0.03101584	0.09470422	0.70208441		

(cont'd on next page)

TABLE 7-3 (cont'd)

n	r_1	r_2	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
10	2	3			0.20496319	0.10382533	0.11220127	0.11982080	0.45918942			
					—0.88982266	—0.11005067	—0.02620385	0.05494874	0.97112842			
	2	4			0.06055627	0.09352928	0.11776906	0.72814539				
					—1.19522527	—0.13182552	—0.01442865	1.34147944				
	2	5			—0.26483993	0.07347216	1.19136777					
					—1.79471038	—0.16877724	1.96348761					
	2	6			—1.34060718	2.34060718						
					—3.56767731	3.56767731						
	3	3				0.38067840	0.11932160	0.11932160	0.38067840			
						—1.28320587	—0.05586351	0.05586351	1.28320587			
	3	4				0.18711959	0.11984110	0.69303931				
						—1.97905791	—0.05532554	2.03438345				
	3	5				—0.48466706	1.48466706					
						—3.95105520	3.95105520					
	4	4					0.50000000	0.50000000				
							—4.07605089	4.07605089				

Reprinted with permission. Copyright © by Institute of Mathematical Statistics.

The reliability of the system is always one minus the cdf of failure times, i.e., one minus the quantity resulting from Eq. 7-17.

Correspondingly, for the series system and different failure distributions for the n components, the overall system reliability still depends on the minimum failure time, or x_1. Therefore, the series system reliability is given by the quantity

$$Pr[x_1 > x] = 1 - \prod_{i=1}^{n} [1 - P_i(x)]. \tag{7-18}$$

In recent years there has been considerable interest and much research on "robust" estimation techniques, and as David (Ref. 16) points out, the order statistics play a very prominent role here because the central observations in an ordered sample are much less liable to be affected by both any spurious observations and the assumptions than are the extreme sample values. As an example, for robust estimation of the population mean, the median and the midmean, or inner 50% sample values, are more robust than the sample average. The sample median is also an example of extreme "trimming" since it involves only the single middle sample value or the average of the two central values.

We have already indicated the idea of "data compression" in effect by the analysis of data as in Example 7-4. In fact, there are many occasions for which one actually will have to deal with large samples, and yet he will not always want to carry out extensive computations with a large mass of data nor will he want to obtain quick estimates. Hence the analyst may desire to "compress" the data or use only a few of the inner ordered sample values. As David (Ref. 16) says, if only two sample order statistics are used to estimate the normal population mean, then from large sample theory such an estimate would be based on the 27th and 73rd percentiles. In other words, the optimal estimate of the normal population mean μ^* for large samples would be

$$\mu^* = [x(0.2708) + x(0.7292)]/2, \tag{7-19}$$

or, in other words, for a sample of $n = 100$, one would take as the optimal estimate of the population mean the quantity

$$\mu^* = (x_{28} + x_{73})/2.$$

David also discusses selection procedures in Ref. 15, in which one is interested in selecting the top k scorers in a certain test taken by n (greater than k) individuals or students, and he gives an example. Another selection procedure might involve just how well individuals selected because of their scores on a test X may be expected to perform on a test Y, say. Like the X scores, Y is also a random variable that presumably may be related to X in a linear fashion. Thus in using the order statistics X_i for the X test scores, there is associated a Y value for the same individual, which we designate by $Y_{[i]}$. This latter sample value for such a bivariate arrangement is known as a "concomitant" of the ith order statistic X_i, so branded by David.

The double sampling scheme discussed by David (Ref. 16) also usually involves a concomitant variable, which is sampled to save time or because tests are expensive or destructive in nature, along with the primary variable of interest. The concomitant variables may also be related to the primary X's through a regression relation.

7-7.2 STATISTICS OF EXTREMES

For many years the primary development of statistical methods lay in the assumption of a normal population or universe, and investigators of the applicable theory directed their attention almost totally toward the axiom of the Gaussian curve. No doubt, much of this may be attributed to the fact that so many problems, for example, in agriculture, demanded immediate solutions, and the analysis of data, perhaps to obtain the best interpretations, had to move forward. However, many investigators, who acquired extensive knowledge with so-called "real world" data, began to note very clearly that often the normal assumption was not too trustworthy, and some even exclaimed that "normality was a myth". In perhaps a large number of applications, it was not easy to disprove normality. For some of the critical, nonnormal problems demanding

extensive analyses, it could be said that national interest had been aroused. One of these critical statistical problems had to do with the problem of floods. Floods, of course, represent extremal values or conditions and occur with very low frequency on a relative basis. One of our former presidents' water commission pointed out that, "However big floods get, there will always be a bigger one coming; so says one theory of extremes, and experience suggests it is true.". Thus for planning flood control projects there is a great deal of interest in the probability distribution of largest values or largest extremes, the distribution of the number of "exceedances" (occurrences equal to or larger than a certain large value), and the expected time interval between floods. Studies of the statistics of extremes were undertaken in a very thorough manner by Gumbel, who published a most comprehensive book on the general subject in 1958 (Ref. 17). Gumbel pointed out (Ref. 17, pp. 21-3) that for the distribution of repeated occurrences and the number of exceedances, one is interested in the probability that the exceedance happens for the first time at a number of trials equal to v, say. Thus the random variable v is an integer, unlimited to the right, and for the event to have happened for the first time at trial v, it must have failed for all of the preceding $(v-1)$ trials. Hence the probability of this is

$$Pr[v] = [F(x)]^{v-1}[1 - F(x)] \tag{7-20}$$

where $F(x)$ is the chance of a value less than a particular (large) observation x.

The mean number of trials to an occurrence or between occurrences, i.e., the "return period" $T = E(v)$, is clearly given by

$$E(v) = T = 1/[1 - F(x)] \tag{7-21}$$

a rather self-evident result. The approximate standard deviation of the number of trials v is (Ref. 17)

$$\sigma(v) = (T^2 - T)^{1/2} \rightarrow T - 1/2 \tag{7-22}$$

so that if $[1 - F(x)]$ is small, indicating a large value of the occurrence x and hence a small upper tail area of the distribution, the return period is very large and the spread of the distribution becomes huge also.

As pointed out by Gumbel (Ref. 17), the cumulative probability that the event happens before or at the vth trial is

$$G(v) = 1 - [F(x)]^v \approx 1 - \exp(-v/T) \tag{7-23}$$

if the return period T is large. (A T greater than, say, 10 or 15 will even give a satisfactory approximation for practical purposes.)

The cumulative probability $G(T)$ for the exceedance to happen at or before the return period T is

$$G(T) = 1 - (1 - 1/T)^T \approx 1 - 1/e = 0.63212. \tag{7-24}$$

Example 7-5:

Given any general, but unknown, distribution of occurrences and some interest in records above the 99% point, or upper 1% tail area chance. Find the expected number of trials to a record or between records, the standard deviation of such a distribution, and the chance that at least 200 trials or observations will be required to reach another exceedance.

In answer to the first question, the return period T is simply

$$E(v) = T = 1/0.01 = 100.$$

Moreover, the standard deviation is, for all intents and purposes, equal to the expected value, or that is, from Eq. 7-22, $100 - 0.5 = 99.5$ for sigma.

The chance that at least 200 trials will be experienced before a record or exceedance is approximately

$$\exp(-200/100) = 0.14.$$

Gumbel's book contains a wealth of information on that phase of order statistics relating to extreme values, including statistical characteristics of extremes for an exponential distribution, the normal distribution, the lognormal distribution, the Cauchy-type distributions, and the Pareto distribution. Asymptotic or large sample characteristics of extreme values are most thoroughly covered. For large samples the largest observation and the smallest observation, or even the ℓth largest and the mth smallest observation, are asymptotically distributed independently (Gumbel, Ref. 17, p. 110). Gumbel also covers the distributional properties of the range of samples and the relation of the range to the problem of tolerance limits of distributions that we discuss briefly in par. 7-7.5. In summary, Gumbel's Ref. 17 represents a book that may be highly useful for many Army applications of the theory of order statistics and extreme values.

7-7.3 GUMBEL'S EXTREME VALUE DISTRIBUTION

A very important and now widely used probability distribution is that of Gumbel (Ref. 17, p. 159); he has characterized it as the Type I asymptotic distribution for the smallest extreme value. Here we will have to limit our discussion for the sake of brevity to taking a rather general form of a "robust" distribution, or model of many different shapes, the well-known and widely used Weibull distribution, and transform it to the Gumbel extreme-value distribution. Let us consider for the moment the two-parameter Weibull time-to-fail probability distribution, for which the chance of observing a failure time T less than t for an item on test is given by

$$Pr[T < t] = F(t) = 1 - \exp[-(t/\theta)^{\beta}],\ t \geq 0 \tag{7-25}$$

where

θ = characteristic life or scale parameter, $\theta > 0$

β = shape parameter, $\beta > 0$.

Now, we transform the time-to-fail variable t and the shape parameter β as follows:

$$X = \ln T, \tag{7-26}$$

$$b = 1/\beta. \tag{7-27}$$

These two transformations, when substituted in Eq. 7-25, yield the new cumulative probability distribution

$$Pr[X < x] = G(x) = 1 - \exp[-\exp\{(x - u)/b\}] \tag{7-28}$$

where

$u = \ln\theta$.

The distribution function (Eq. 7-28) is widely known as Gumbel's extreme-value distribution, and it is seen that if one studies the properties of the extreme-value distribution, he can also make inferences about the original two-parameter Weibull distribution. In fact, this is precisely what has been done by many investigators delving into the theory of reliability and life testing. In this connection and as a source of some examples, we suggest that the reader might consult Mann, Schafer, and Singpurwalla's book on methods for the statistical analysis of reliability and life data (Ref. 18). Many uses are given there, as are also indicated by Gumbel (Ref. 17).

Incidentally, the reader will note that for the original Weibull law of Eq. 7-25, the scale parameter is transformed to a "location" parameter in Eq. 7-28, and the shape parameter becomes a "scale" parameter.

As the sample size n increases, the least and greatest sample values, or the "extremes", and even the ℓth largest and mth smallest values will approach limiting distributions. Thus when the extremes are transformed or otherwise standardized, they will approach a limiting distribution, which for a wide class of distributions converge to only about three types, including the Gumbel least extreme and greatest extreme value distributions. In effect, therefore, we have an important class of parent populations, including the normal distribution (we illustrated only the Weibull), for which the limiting distribution is the doubly exponential extreme-value distributions. See Refs. 4, 5, 17, and 18 for details.

7-7.4 ORDER STATISTICS AND OUTLYING OBSERVATIONS

By referring to Chapter 3 of this handbook, it is seen that tests for outliers or discrepant values in a sample almost invariably turn out to be significance tests for certain of the sample order statistics, especially the largest and/or smallest few observations. Indeed, consider the largest, extreme residual Studentized ratio given by Eq. 3-32 or the Studentized deviation from the sample mean of the smallest observation in Eq. 3-34. These Studentized ratios involve the first and nth order statistics of the sample.

The Studentized range of Eq. 3-37 or Eq. 7-2 is based on the least and greatest sample values, or the first and nth order statistics, and in a significance test they would be used to judge whether the sample extremes are too far apart, i.e., whether x_1 and x_n are simultaneously outliers, perhaps. As was seen in Chapter 3, however, this may not be a completely satisfactory test, for either or both of the sample extremes could be outliers. On the other hand, if faced with such a situation, we could ignore the least and greatest sample values of the sample and use the remaining order statistics to estimate, for example, the population mean and standard deviation with quite acceptable efficiency. That is to say, we could censor x_1 and x_n from consideration and use the order statistic approach instead. See Example 7-6.

The Studentized extreme deviate tests, the Studentized range, the Dixon sample criteria of par. 3-5.2, the Tietjen-Moore tests of par. 3-5.5.2, the Rosner and Hawkins multiple outlier detection procedures of par. 3-5.5.3, and other outlier screening procedures of Chapter 3 all depend in some way on the use of specific order statistics or significance tests. In fact, the outlier detection techniques should be quite sensitive to shifts in level or scale for many of the sample observations, so that aberrant values will be branded. However, it is usually such shifts in level or scale that lead to nonhomogeneous or nonrepresentative samples drawn from some population(s) of which we are trying to learn the properties. Thus the aberrant sample values or outliers will place our estimate of the population mean in the wrong position, or they will inflate the estimate of the population standard deviation, etc., thereby leading to nonrobust or poor estimators fraught with biases. In fact, it is interesting to consider again the data of Example 3-5 for the 15 vertical semidiameter measurements of the planet Venus.

Example 7-6:

Return to the data of Example 3-5 and reconsider the decision to reject the least sample value of -1.40 and the largest value of 1.01 especially since the Tietjen-Moore tests rejected both values, the Rosner test did not, and the Hawkins test found the two values to be significant. Since we now may use an order statistic analysis to estimate the normal population mean and sigma, we can compare estimators for the original sample, the remaining sample after rejection of the values -1.40 and 1.01, and the estimates of the universe mean and sigma based on the use of sample order statistics x_2 through x_{n-1}.

For the original sample of 15 observations, the mean $\bar{x}$ and standard deviation s are

$$\bar{x} = 0.018 \text{ and } s = 0.551$$

so that perhaps we could be disturbed by the size of s. Hence if we were to reject the "outliers" -1.40 and 1.01 and then determine a new mean and sigma from the remaining 13 observations of the sample, we would get

$$\bar{x} = 0.051 \text{ and } s = 0.322$$

which gives an increase of 183% in mean value and a decrease of 42% in the standard deviation! Finally, if we were to estimate the normal population mean and sigma by using the sample order statistics x_2 through x_{14}, and thus censor the -1.40 and 1.01 from any consideration, our estimated mean and standard deviation become

$$\bar{x} = 0.056 \text{ and } s = 0.427.*$$

*For the sample of size $n = 15$ and x_1 and x_{15} censored, the coefficients to calculate the mean and sigma were taken from Table 10C.1, p. 232, of Sarhan and Greenberg's book (Ref. 4).

We note in this connection that the trimmed and censored samples give equal estimates of the population mean but that the trimmed sample gives a smaller sigma (0.322) than does the censored sample (0.427). Obviously, it cannot be said that the trimmed sample would give an unbiased estimate of the scale parameter although it might be expected to give an unbiased mean. On the other hand, the censored sample does indeed give unbiased, minimum variance estimates of both the population mean and sigma even though the least and greatest sample values were not included. Therefore, just in case something may have happened to the sample values, one would tend to place more confidence in the censored sample theory and to take 0.056 as the mean and 0.427 as the proper sigma.

7-7.5 UNIVARIATE TOLERANCE INTERVALS

Whereas many applications of statistical methods call for the estimation of population parameters and the determination of confidence bounds on the true unknown parameters, such as the mean and standard deviation, another very useful, and often more important, problem is that of estimating with high confidence the fraction or percentage of a population (distribution) within two limits or bounds. For the ordered sample statistics, for example, it seems natural to estimate the fraction of the population sampled between the highest and lowest values of the sample, i.e., the use of the range as a "tolerance limit". In this connection, we have that the cumulative probability up to the least sample value x_1 is $F(x_1)$, and the cumulative population probability up through the largest sample value x_n is $F(x_n)$. Hence the difference $[F(x_n) - F(x_1)]$ is actually the fraction of the sampled population bounded by the sample range. Therefore, we might consider two functions of the sample values—such as the end points of the range x_1 and x_n or the two sample order statistics x_r and x_s with $1 \leq r \leq s \leq n$—and try to discover just what probability statements can be made about the fraction of the sampled population between such limits. This type of statistical problem was studied initially by Wilks (Ref. 19) who showed that, for any fraction γ of the population between the range limits and confidence β, the following probability statement holds

$$\begin{aligned} \beta = Pr\{[F(x_n) - F(x_1)] \geq \gamma\} &= 1 - I_\gamma(n-1,2)^* \quad [\text{or } \beta = I_{1-\gamma}(2,n-1)] \\ &= \sum_{i=0}^{n-2} \binom{n}{i} \gamma^i (1-\gamma)^{n-i} * \end{aligned} \tag{7-29}$$

where we see that the chance of including various fractions of the sampled population between range limits can be expressed in terms of the incomplete beta function ratio or a binomial sum. In fact, Wilks (Ref. 19) also showed that for x_r and x_s, we have

$$\begin{aligned} \beta = Pr\{[F(x_s) - F(x_r)] \geq \gamma\} &= 1 - I_\gamma(s-r, n-s+r+1) \ , \ r < s \\ &= \sum_{i=0}^{s-r-1} \binom{n}{i} \gamma^i (1-\gamma)^{n-i}. \end{aligned} \tag{7-30}$$

Wilks' results (Ref. 19) amount to a very fine accomplishment or "breakthrough" indeed because they establish that no matter what the distributional form of the continuous population sampled, one can nevertheless make a probability or confidence statement about the fraction of the population that is included between either the range limits of the sample or between any two sample order statistics! Alternatively, one may determine in advance the sample size required to guarantee that at least a certain fraction of the population will be included between the sample range limits with a given degree of assurance . Thus it is for such reasons that Wilks' results (Ref. 19) are referred to as "distribution-free tolerance limits". In fact, before this result was obtained, one usually had to be content with just placing confidence bounds on each parameter of some assumed distribution. Finally, the population tolerance interval statements covered by Eqs. 7-29 and 7-30 turn out to be very simple mathematically.

---

*For any continuous general distribution, the central area $W = F(x_n) - F(x_1)$ has a pdf $g(W) = n(n-1)W^{n-2}(1-W)$.

For the tolerance interval covered by the sample range limits (x_1, x_n), it is easily seen that the last RHS of Eq. 7-29 reduces to a very simple relation between the confidence level or probability β, the fraction of the population covered by the range limits or γ, and the sample size n. This simple relation for any continuous distribution is

$$\beta = 1 - n\gamma^{n-1} + (n-1)\gamma^n. \tag{7-31}$$

Thus if we know any two of the parameters, the other or unknown value may be found with the sample size n by cut-and-try or iteration, or building a table.

Eqs. 7-29 and 7-30 may be evaluated by using AMCP 706-109, *Tables of the Cumulative Binomial Probabilities* (Ref. 20), i.e., by the relations

$$\beta = P(2, n, 1-\gamma) \quad \text{for Eq. 7-29} \tag{7-32}$$

and

$$\beta = P(n-s+r+1, n, 1-\gamma) \quad \text{for Eq. 7-30.} \tag{7-33}$$

In fact, it is very easy to use the tables of Ref. 20 for numerous such calculations if desired.

In the statistical literature there are some graphs and tables the analyst may use to advantage concerning the applied problems of "distribution-free" or "nonparametric" tolerance limits. Gumbel (Ref. 17) on his Graph 3.2.4 gives the relation among the sample size, the confidence level, and the fraction of the population *outside* the sample range limits. Gumbel uses a logarithmic scale for the sample size and the fraction *outside* range limits so that the confidence or probability curves are straight lines.

Murphy (Ref. 21) gives three useful graphs—one for each of the confidence levels of 90%, 95%, and 99%—and the corresponding relations between the amounts of population "coverage", the sample size, and the number m of intervals or "blocks", which are excluded from tolerance region runs. The term coverage is used to define the amount or fraction of the population sampled between any two order statistics. For example, for the sample range the fraction of population coverage would be $[F(x_n) - F(x_1)]$, etc. With regard to the definition of the term "block", we first think of the n sample order statistics as being plotted along the x-axis so that the sample space is then divided into $(n+1)$ intervals or blocks. Therefore, it can be said that the term block has been used to extend or generalize this concept to two or more dimensions (Murphy, Ref. 21). Now, if we think of r as referring to the rth smallest sample order statistic x_r and s as referring to the sth largest order statistic x_s, the pdf for the central area of the distribution W given by

$$W = F(x_{n-s+1}) - F(x_r) \tag{7-34}$$

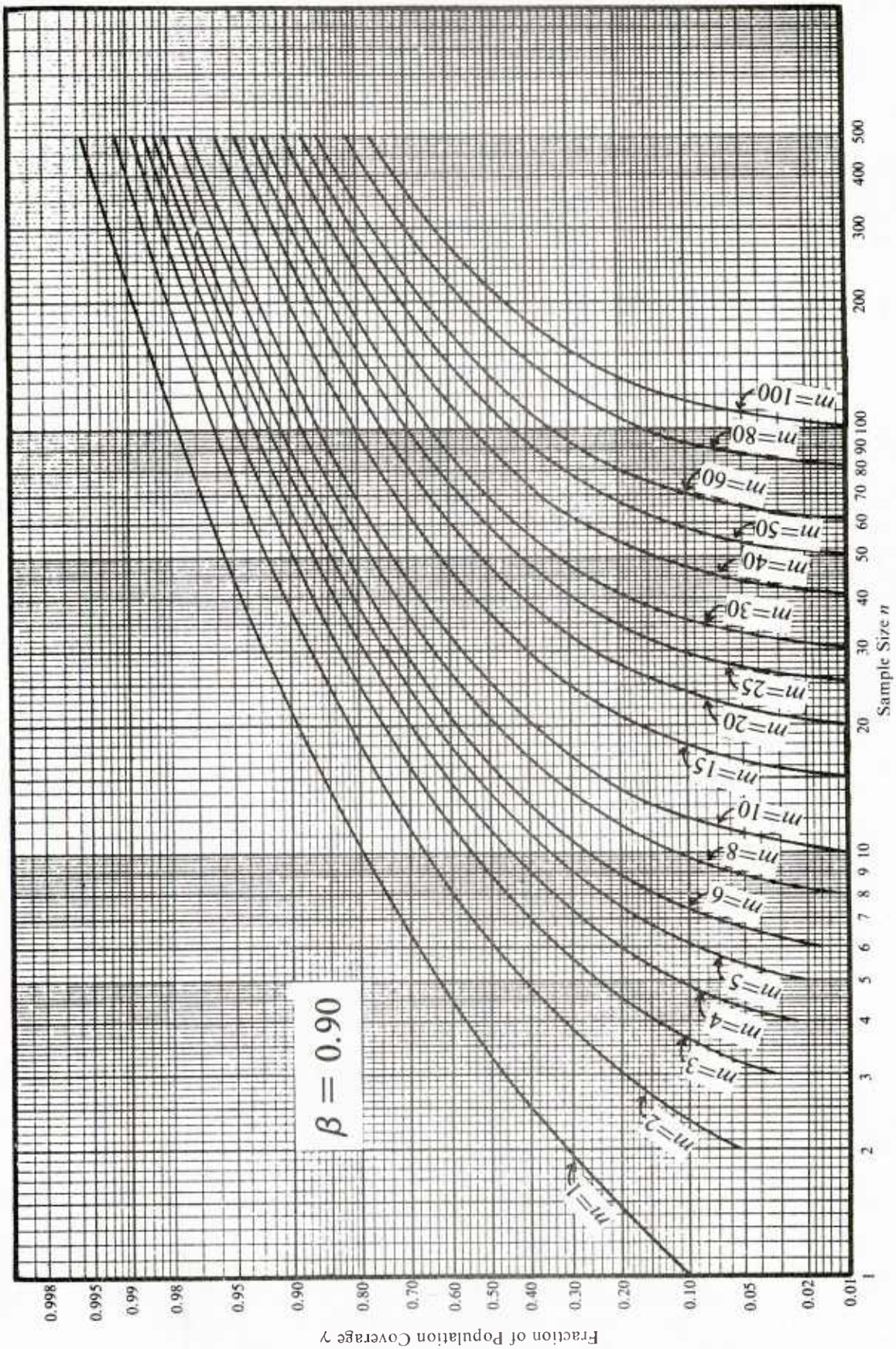
is

$$g(W) = \frac{\Gamma(n+1)}{\Gamma(n-m+1)\,\Gamma(m)} W^{n-m}(1-W)^{m-1} \tag{7-35}$$

where

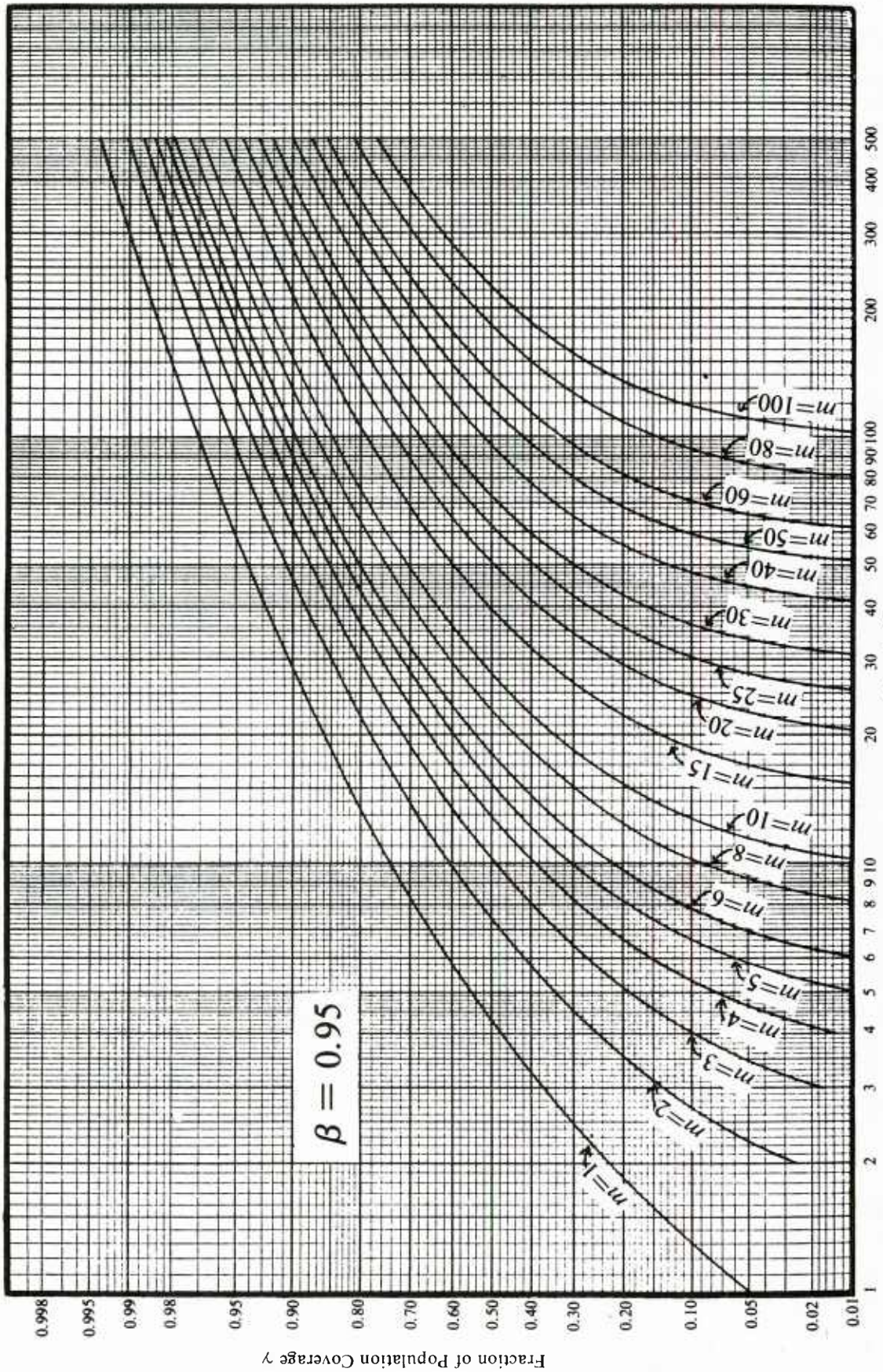
$$\Gamma(\,) = \text{complete gamma function of quantity in parentheses} \tag{7-36}$$
$$m = r + s.$$

Thus we see that m is the total number of blocks below the rth smallest and above the sth largest observations that are *excluded*. For the sample range, therefore, $m = 2$, and if we deal with the next to the largest and next to the smallest values, we would have $m = 4$, etc. Murphy (Ref. 21) gives some graphs of the coverage (Eq. 7-34) on his Figs. 1, 2, and 3, which we reproduce as Figs. 7-1, 7-2, and 7-3. Note that the sample sizes run from $n = 1$ to 500; there are three confidence, probability, or "tolerance" levels of β, i.e., $\beta = 0.90, 0.95$, and 0.99; and the ordinate of each figure is the fraction of population coverage γ. The number m of excluded intervals or blocks runs from the curve $m = 1$ for the sample range end points to $m = 100$—a very wide coverage indeed!



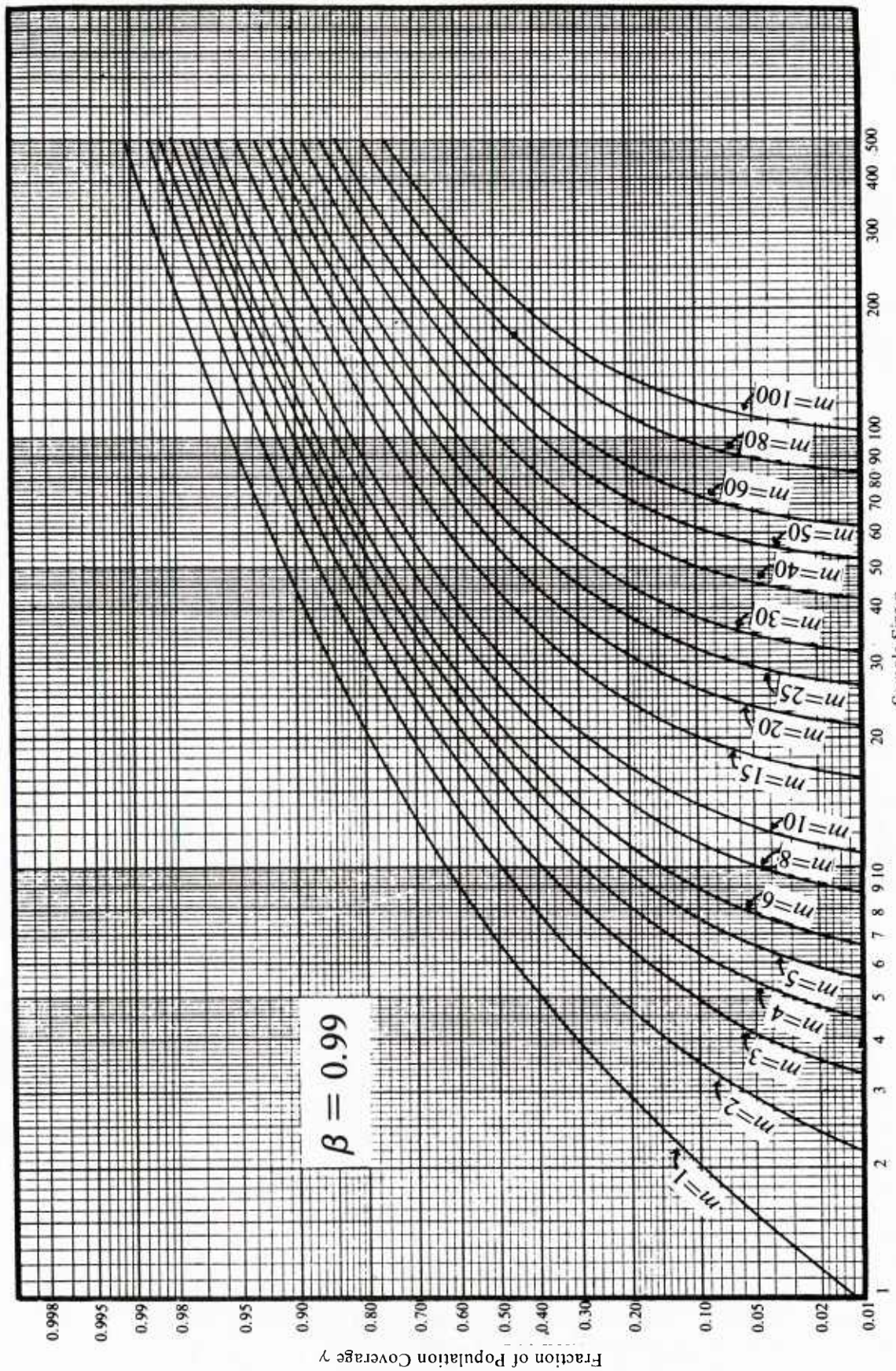
Reprinted with permission. Copyright © by Institute of Mathematical Statistics.

Figure 7-1. Graphs of Population Coverage for the Tolerance Level $\beta = 0.90$ (Ref. 21)



Reprinted with permission. Copyright© by Institute of Mathematical Statistics.

Figure 7-2. Graphs of Population Coverage for the Tolerance Level $\beta = 0.95$ (Ref. 21)



Reprinted with permission. Copyright © by Institute of Mathematical Statistics.

Figure 7-3. Graphs of Population Coverage for the Tolerance Level $\beta = 0.99$ (Ref. 21)

In addition to the curves of Murphy (Ref. 21) for the tolerance limits problem, Somerville (Ref. 22) later published two very useful tables that are quite compact and hence are included here as Tables 7-4 and 7-5. Table 7-4 is for fractional population coverages, which are fixed at 0.50, 0.75, 0.90, 0.95, and 0.99, for confidence levels β equal to precisely the same values as for γ, for sample sizes n = 50(5)100(10)150(20)170(30)200(100)1000 and with $m = r + s$ given by the values listed in the body of Table 7-4. In using the sample range limits, for example, one would select only those particular values of m within the table that are listed as $m = 2$.

Within the body of the table, Table 7-5 gives the values of the confidence or probability β that will guarantee at least the amount of population coverage $\gamma = 0.50, 0.75, 0.90, 0.95$, or 0.99 and for sample sizes n = 3(1)15(2)17(1)20(5)30(10)100. Example 7-7 illustrates the use of the referenced figures and tables.

Example 7-7:

Given a sample of size 25, which has been selected at random from a population believed to be a gamma distribution with perhaps a rather long tail to the right. Without estimating any parameters of the population, it is very important to know just how much of the sampled universe could be included within sample range limits with 90% assurance. How large a sample would be necessary to state with 90% assurance that at least 95% of the sampled population would be included within range limits?

Of course, tolerance limits may be determined no matter what type of population is sampled, provided it is continuous—a reasonable assumption in this case. For the answer to the first question, it is easily seen by examining Fig. 7-1 that for $n = 25$ and the $m = 2$ curve, one can state with 90% assurance that at least about 85% of the population would fall within range limits.

To answer the second question, one may examine Fig. 7-1 for 90% assurance and note that the curve for $m = 2$ intersects the 95% coverage line of a distribution at about $n = 77$. Moreover, a look at Table 7-5 for γ = 95% will show that a sample size of $n = 80$ will provide 91% assurance that the sample range limits will cover at least 95% of the sampled population. Hence just a trifle under $n = 80$ is needed, so that about $n = 78$ would be sufficient. (If desired, one could nearly infer this result from Table 7-4.)

It is perhaps of some further interest to this example that we add the additional knowledge which states that if one desires to cover 99% of the population with the observed sample range end points and also with 90% confidence, a sample of size $n = 400$ would be required (Table 7-4), indicating the "cost" in terms of sample size.

With regard to the use of Wilks' tolerance limits for general populations, a very natural and important question to ask would be, "What amount of information is lost or what 'inflated' sample size is suffered, due to the 'robust' assumption of sampling any 'continuous distribution'?". Thus if one knows quite well the type of population he is sampling, cannot a justifiable gain in information or decrease in sample size be attained? The answer to such a question is very decidedly "yes"—quite an increase in information or a decrease in sample size can be achieved. In fact, there can also be quite a gain in flexibility because for the normal distribution, for example, for just about any sample size one can provide confidence limits based on the sample mean and standard deviation, or the sample range, which will include at least some fraction of the normal population for future samples and also for any given level of confidence. This particular problem for sampling a normal universe has been studied by, for example, Bowker (Refs. 23 and 24), who used the sample standard deviation, and by Mitra (Ref. 25), who used the sample range instead of the sample sigma. In view of the simplicity of the sample range and the fact that we have used it previously in connection with Wilks' tolerance limits for general distributions, we will limit our discussion to that sample statistic, i.e., its end points. Table 7-6 is taken from the paper of Mitra (Ref. 25) and gives, in the body of the table, values of k for which tolerance limits based on $(\bar{x} - kw)$ and $(\bar{x} + kw)$ for the sampled normal population will include at least $\gamma = 0.75, 0.90, 0.95$, or 0.99 of the normal universe with confidence levels of $\beta = 0.75, 0.90, 0.95$, or 0.99. Note in particular and especially for small sample sizes that the distance between tolerance limits in Table 7-6 can be very wide indeed, whereas in our account of Wilks' general distribution tolerance limits, the bounds are the sample range. Thus for comparative purposes one would have to attain a value of k that equals one-half in order to have the same width limits. Nevertheless, we recall from Example 7-7 that a sample size of $n = 77$ would be required to give sample range end points that would cover 95% of the general population with 90% assurance.

TABLE 7-4

VALUES OF $m = r + s$ SUCH THAT WE MAY ASSERT WITH CONFIDENCE AT LEAST β THAT 100_γ PERCENT OF A POPULATION LIES BETWEEN THE rth SMALLEST AND THE sth LARGEST OF A RANDOM SAMPLE OF n FROM THAT POPULATION (CONTINUOUS DISTRIBUTION FUNCTION ASSUMED) (Ref. 22)

n	γ																								
	$\beta = 0.50$					$\beta = 0.75$					$\beta = 0.90$					$\beta = 0.95$					$\beta = 0.99$				
	0.50	0.75	0.90	0.95	0.99	0.50	0.75	0.90	0.95	0.99	0.50	0.75	0.90	0.95	0.99	0.50	0.75	0.90	0.95	0.99	0.50	0.75	0.90	0.95	0.99 = γ
50	25	12	5	2	0	22	10	3	1	—	20	9	2	1	—	19	8	2	—	—	16	6	1	—	—
55	28	14	5	3	0	25	12	4	2	—	23	10	3	1	—	21	9	2	—	—	19	7	1	—	—
60	30	15	6	3	0	27	13	4	2	—	25	11	3	1	—	24	10	2	1	—	21	8	1	—	—
65	33	16	6	3	0	30	14	5	2	—	27	12	4	1	—	26	11	3	1	—	23	9	2	—	—
70	35	17	7	3	1	32	15	5	2	—	30	13	4	1	—	28	12	3	1	—	25	10	2	—	—
75	38	19	7	4	1	35	16	6	2	—	32	14	4	1	—	30	13	3	1	—	27	10	2	—	—
80	40	20	8	4	1	37	17	6	3	—	34	15	5	2	—	33	14	4	1	—	30	11	2	—	—
85	43	21	8	4	1	39	19	7	3	—	37	16	5	2	—	35	15	4	1	—	32	12	3	—	—
90	45	22	9	4	1	42	20	7	3	—	39	17	5	2	—	37	16	5	1	—	34	13	3	1	—
95	48	24	9	5	1	44	21	7	3	—	41	18	6	2	—	39	17	5	2	—	36	14	3	1	—
100	50	25	10	5	1	47	22	8	3	—	44	20	6	2	—	42	18	5	2	—	38	15	4	1	—
110	55	27	11	5	1	51	24	9	4	—	48	22	7	3	—	46	20	6	2	—	43	17	4	1	—
120	60	30	12	6	1	56	27	10	4	—	53	24	8	3	—	51	22	7	2	—	47	19	5	1	—
130	65	32	13	6	1	61	29	11	5	—	58	26	9	3	—	56	25	8	3	—	52	21	6	2	—
140	70	35	14	7	1	66	31	12	5	1	62	28	10	4	—	60	27	8	3	—	56	23	6	2	—
150	75	37	15	7	1	71	34	12	6	1	67	31	10	4	—	65	29	9	3	—	61	26	7	2	—
170	85	42	17	8	2	81	39	14	7	1	77	35	12	5	—	74	33	11	4	—	70	30	9	3	—
200	100	50	20	10	2	95	46	17	8	1	91	42	15	6	—	88	40	13	5	—	84	36	11	4	—
300	150	75	30	15	3	144	70	26	12	2	139	65	23	10	1	136	63	22	9	1	130	58	19	7	—
400	200	100	40	20	4	193	94	36	17	3	187	89	32	15	2	184	86	30	13	1	177	80	27	11	—
500	250	125	50	25	5	242	118	45	22	3	236	113	41	19	2	232	109	39	17	2	224	103	35	14	1
600	300	150	60	30	6	292	143	55	26	4	284	136	51	23	3	280	133	48	21	2	272	126	44	18	1
700	350	175	70	35	7	341	167	65	31	5	333	160	60	28	4	328	156	57	26	3	319	149	52	22	2
800	400	200	80	40	8	390	192	74	36	6	382	184	69	32	5	377	180	66	30	4	367	172	61	26	2
900	450	225	90	45	9	440	216	84	41	7	431	208	79	37	5	425	204	75	35	4	415	195	70	30	3
1000	500	250	100	50	10	489	241	94	45	8	480	233	88	41	6	474	228	85	39	5	463	219	79	35	3

Reprinted with permission. Copyright © by Institute of Mathematical Statistics.

TABLE 7-5

CONFIDENCE β WITH WHICH WE MAY ASSERT THAT 100 γ PERCENT OF THE POPULATION LIES BETWEEN THE sth LARGEST AND rth SMALLEST OF A RANDOM SAMPLE OF n FROM THAT POPULATION (CONTINUOUS DISTRIBUTION ASSUMED) (Ref. 22)

n	γ=0.50	γ=0.75	γ=0.90	γ=0.95	γ=0.99	n	γ=0.75	γ=0.90	γ=0.95	γ=0.99
3	0.50	0.16	0.03	0.01	0.00	17	0.95	0.52	0.21	0.01
4	0.69	0.26	0.05	0.01	0.00	18	0.96	0.55	0.22	0.01
5	0.81	0.37	0.08	0.02	0.00	19	0.97	0.58	0.25	0.02
6	0.89	0.47	0.11	0.03	0.00	20	0.98	0.61	0.26	0.02
7	0.94	0.56	0.15	0.04	0.00	25	0.99	0.73	0.36	0.03
8	0.96	0.63	0.19	0.06	0.00	30	1.00	0.82	0.45	0.04
9	0.98	0.70	0.23	0.07	0.00	40		0.92	0.60	0.06
10	0.99	0.76	0.26	0.09	0.00	50		0.97	0.72	0.09
11	0.99	0.80	0.30	0.10	0.01	60		0.99	0.81	0.12
12	1.00-	0.84	0.34	0.12	0.01	70		0.99	0.87	0.16
13		0.87	0.38	0.14	0.01	80		1.00-	0.91	0.19
14		0.90	0.42	0.15	0.01	90			0.94	0.23
15		0.92	0.45	0.17	0.01	100			0.96	0.26
		0.94	0.49	0.19	0.01					

x_r = rth smallest sample observation
x_s = sth largest sample observation

Reprinted with permission. Copyright © by Institute of Mathematical Statistics.

Clearly, the value of the sample size sought is well beyond the highest one, $n = 20$, given in Table 7-6 for the middle column of the section for $\beta = 0.95$. However, one can, by cut-and-try methods, use jointly Eqs. 2.1 and 2.2 of Mitra's paper (Ref. 25) along with Harter's tables of the percentage points of the range (Ref. 1, p. 374) to see that a sample size of no more than about $n = 50$ is required when it is known that the population sampled is indeed a normal universe. Thus it can be said that exact knowledge of the particular form of the population does save very significantly insofar as the sample size "cost" is concerned. Similar computations would further clarify the general subject and no doubt would have practical value.

Although in our account of tolerance intervals for the normal population, we have used only the sample range by way of illustration, we should point out that Bowker (Ref. 23, pp. 102-7, Table 2.1) gives very extensive coverage for the use of the sample standard deviation. In fact, his sample sizes go up through the value $n = 1000$, and a reference line for $n = \infty$ also is included at the bottom of the table. Therefore, we recommend use of Bowker's Table 2.1 as practical applications demand.

We have covered only the use of univariate tolerance intervals although for some applications the analyst might have the need to apply multivariate tolerance intervals. For such applications see Sarhan and Greenberg (Ref. 4, p. 141) or Murphy (Ref. 21).

Finally, another important use of tolerance intervals relates to the determination of confidence intervals for the various percentage points of distributions—see, for example, Sarhan and Greenberg (Ref. 4, p. 137). For such applications confidence intervals for the lower percentage points are based on the least sample value and some rth smallest observation, and confidence intervals for the upper percentage points use the largest sample observation and the sth largest one (see Ref. 4). As is well-known, the percentage points of distributions are often referred to as "quantiles".

TABLE 7-6

TOLERANCE FACTORS FOR NORMAL DISTRIBUTIONS (Ref. 25)

Factors k such that the probability is β that at lease a proportion γ of the distribution will be included between $\bar{x} \pm kw$ where $\bar{x}$ is the mean and w is the range in a sample of size n.

	$\beta = 0.75$					$\beta = 0.90$				
n \ γ	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
2	3.181	4.456	5.243	6.740	8.429	8.065	11.298	13.294	17.090	21.374
3	1.312	1.857	2.197	2.850	3.591	2.169	3.069	3.631	4.711	5.936
4	0.916	1.301	1.544	2.012	2.546	1.321	1.877	2.227	2.902	3.672
5	0.744	1.060	1.259	1.644	2.086	1.003	1.428	1.697	2.216	2.812
6	0.647	0.923	1.097	1.435	1.824	0.837	1.194	1.420	1.857	2.360
7	0.584	0.834	0.992	1.299	1.652	0.735	1.050	1.248	1.635	2.080
8	0.540	0.771	0.917	1.202	1.530	0.666	0.951	1.131	1.483	1.888
9	0.507	0.723	0.861	1.129	1.438	0.615	0.879	1.046	1.372	1.747
10	0.481	0.687	0.817	1.072	1.366	0.577	0.824	0.981	1.286	1.639
11	0.460	0.657	0.782	1.026	1.308	0.546	0.780	0.929	1.219	1.554
12	0.442	0.632	0.753	0.988	1.260	0.521	0.745	0.887	1.164	1.484
13	0.428	0.611	0.728	0.956	1.219	0.501	0.715	0.852	1.118	1.426
14	0.415	0.594	0.707	0.928	1.184	0.483	0.690	0.822	1.079	1.377
15	0.405	0.578	0.689	0.904	1.154	0.408	0.669	0.797	1.046	1.334
16	0.395	0.565	0.673	0.883	1.127	0.455	0.650	0.774	1.016	1.297
17	0.386	0.553	0.658	0.864	1.103	0.443	0.633	0.755	0.991	1.265
18	0.379	0.542	0.645	0.848	1.082	0.433	0.619	0.737	0.968	1.235
19	0.372	0.532	0.634	0.833	1.063	0.424	0.605	0.721	0.947	1.209
20	0.366	0.523	0.623	0.819	1.045	0.415	0.594	0.707	0.929	1.186

	$\beta = 0.95$					$\beta = 0.99$				
n \ γ	0.75	0.90	0.95	0.99	0.999	0.75	0.90	0.95	0.99	0.999
2	16.158	22.635	26.634	34.238	42.821	80.972	113.429	133.469	171.576	214.588
3	3.109	4.399	5.206	6.752	8.509	7.034	9.951	11.776	15.275	19.249
4	1.704	2.422	2.873	3.744	4.737	2.978	4.233	5.021	6.543	8.279
5	1.228	1.749	2.078	2.715	3.444	1.903	2.709	3.219	4.205	5.335
6	0.995	1.418	1.686	2.206	2.803	1.433	2.042	2.429	3.178	4.038
7	0.856	1.222	1.453	1.903	2.420	1.176	1.678	1.996	2.615	3.325
8	0.764	1.090	1.297	1.700	2.165	1.015	1.449	1.724	2.261	2.878
9	0.698	0.997	1.187	1.556	1.981	0.903	1.290	1.536	2.014	2.565
10	0.648	0.926	1.103	1.446	1.843	0.823	1.176	1.400	1.836	2.340
11	0.610	0.871	1.037	1.361	1.735	0.762	1.088	1.296	1.701	2.168
12	0.578	0.827	0.985	1.292	1.648	0.714	1.020	1.215	1.594	2.033
13	0.553	0.790	0.940	1.235	1.575	0.675	0.964	1.148	1.507	1.922
14	0.531	0.759	0.904	1.187	1.514	0.642	0.917	1.093	1.435	1.830
15	0.513	0.733	0.873	1.146	1.462	0.614	0.878	1.046	1.373	1.753
16	0.497	0.710	0.845	1.110	1.417	0.591	0.845	1.007	1.322	1.687
17	0.482	0.690	0.822	1.109	1.377	0.571	0.816	0.972	1.277	1.630
18	0.470	0.672	0.801	1.051	1.342	0.553	0.790	0.941	1.236	1.578
19	0.459	0.656	0.782	1.027	1.311	0.538	0.768	0.916	1.203	1.535
20	0.449	0.642	0.765	1.005	1.282	0.524	0.743	0.892	1.171	1.495

Reprinted with permission. Copyright © by the American Statistical Association.

7-8 ORDER STATISTICS AND THE RELATED FIELDS OF RELIABILITY AND LIFE TESTING

Perhaps it could be said that one of the most important uses of sample order statistics is their unique application to the fields of reliability and life testing. We have mentioned that the extensive applications of order statistics to the Army's problems in system reliability and the life testing of items represent very major activities, and we cannot delve into them profoundly in this particular chapter. Nevertheless, we do point out that for the practicing Army analyst, a rather large number of important topics on the use of order statistics in connection with reliability of systems and confidence intervals on system reliability, life testing of items, reliability growth concepts, the availability of military systems to start a mission, and the maintainability of systems are covered in Chapter 21 of Ref. 26.

The two primary probability distributions employed in Chapter 21 of Ref. 26 are the exponential and the Weibull distributions, and sample order statistics are used extensively with both assumptions. For the purposes of this chapter and handbook, there are one or two particular concepts we will review and highlight. These relate to the exponential distribution for time-to-fail type data. We start with the definition of the exponential time-to-fail pdf, which is

$$f(t) = \lambda\exp(-\lambda t) = (1/\theta)\exp(-t/\theta) \tag{7-37}$$

where

$\lambda = 1/\theta$ = failure rate

θ = mean time to fail for the items.

The cdf for Eq. 7-37 is

$$F(t) = 1 - \exp(-t/\theta). \tag{7-38}$$

The exponential distribution (Eq. 7-37 or Eq. 7-38) has a mean value $= 1/\lambda = \theta$, a variance $= 1/\lambda^2 = \theta^2$, with a skewness coefficient of 2 and a kurtosis parameter of 9.

The times to fail of n items, components, systems, etc., placed on test or put into service may be listed as

$$t_1 \leq t_2 \leq t_3 \leq \cdots \leq t_r \leq \cdots \leq t_n$$

where, as indicated, the testing of items may be truncated at the rth failure; otherwise the test could be truncated at a preset or required time.

For the test that is truncated at the rth failure time t_r (or even continued to $r = n$), the ML, minimum variance, unbiased estimator $\hat{\theta}$ of the mean time to fail θ is

$$\hat{\theta} = [\sum_{i=1}^{r} t_i + (n-r)t_r]/r. \tag{7-39}$$

One notes that when $r = n$ in Eq. 7-39, the estimate of the mean time to fail becomes the "usual" one, or

$$\hat{\theta} = \sum_{i=1}^{n} t_i/n. \tag{7-40}$$

If both sides of Eq. 7-39 are multiplied by r, the result $r\hat{\theta}$ is known as the "total time on test" and is a key concept or characteristic in life testing.

The quantity

$$2r\hat{\theta}/\theta = \chi^2(2r) \tag{7-41}$$

i.e., the quantity follows the chi-square distribution with $2r$ df so that confidence bounds are easily placed on the unknown mean time-to-fail parameter θ or on the reliability for some mission time.

Adjacent time differences are independent, and each quantity

$$2(n-r+1)\,(t_r - t_{r-1})/\theta = \chi^2(2), \quad r = 2,\ldots, n \tag{7-42}$$

follows the chi-square distribution with 2 df, i.e., all have an exponential distribution.

The rth ordered random time t_r is the waiting time to obtain the rth failure, and its mean and variance are, respectively,

$$E(t_r) = \theta \sum_{i=1}^{r} [1/(n-i+1)] \tag{7-43}$$

and

$$\mathrm{Var}(t_r) = \theta^2 \sum_{i=1}^{r} [1/(n-i+1)]^2 \tag{7-44}$$

so that the approximate chi-square distribution of par. 4-4.5 may be fitted, or better still, the exact distribution given by Epstein and Sobel (Ref. 27).

Other details of interest may be found in Chapter 21 of Ref. 26, for example, or in Mann, Schafer, and Singpurwalla (Ref. 18). A very complete account of the Weibull distribution may be found in Ref. 18.

For time truncation instead of truncation at a preset number of failures r, one can consider that in effect r failures have taken place at the truncation time t_0 and hence that the ML estimate of the mean time to fail is simply $\hat{\theta} = nt_0/r$. (See, for example, Ref. 18.)

A number of examples using these principles can be found in Chapter 21 of Ref. 26. In particular, we recommend that interested readers review Examples 21-8 and 21-9 of Chapter 21, Ref. 26, and Example 21-10, which applies to the two-parameter negative exponential distribution.

This gives sufficient background for us to turn to the idea of using order statistics in connection with target firings and analyses as they are of much value for such problems.

7-9 THE RADIAL ORDER STATISTICS AND THEIR APPLICATIONS TO TARGET ANALYSES

The analyses of target firings represent some of the prime Army uses of statistics, especially the need for sample order statistics. In such fields of application of statistical methods, we are either dealing with the fall of shot or impacts on the ground; otherwise often we have the problem of analyzing the two-way distribution of impact points or holes from a test involving firings at a vertical target. Moreover, it invariably happens that some of the shots will miss the target; therefore, this problem complicates the statistical analyses of estimation of the parameters of the overall two-dimensional distribution. We usually assume the bivariate normal distribution for impacts. Moreover, in the sequel we will assume that the pattern of shots is "circular", i.e., the standard deviations in the x- and y-directions are equal since this assumption applies to a very large number of target firings—e.g., rifles, many rockets, and other weapons—but not artillery for which the pattern elongates in the range direction. Hence the bivariate normal distribution describing the impact points will have the density

$$f(x,y) = [1/(2\pi\sigma^2)] \exp[-(x^2+y^2)/(2\sigma^2)] \tag{7-45}$$

where x and y are the horizontal and vertical directions, or range and deflection directions, respectively, and each ranges over infinite limits.

If we take the radial distance to any general impact assuming the center of impact is at the origin, we are dealing with the (radial) error r or

$$r = \sqrt{x^2+y^2} \tag{7-46}$$

for which the radial distance r ranges over the limits of zero to infinity, i.e., only positive values. Hence if one applies the usual polar transformation to Eq. 7-45 and integrates out the angular variable, it is well-known that the result is the radial density given by

$$f(r) = (r/\sigma^2)\exp[-r^2/(2\sigma^2)]. \tag{7-47}$$

The density represented by Eq. 7-47 often is referred to as the Rayleigh pdf although it is clearly the chi-square density function with 2 df. We also know that if one sets

$$r^2/2 = t \tag{7-48}$$

then Eq. 7-47 becomes the well-known exponential density function

$$f(t) = (1/\sigma^2)\exp(-t/\sigma^2).* \tag{7-49}$$

One sees, therefore, that if the C of I of the rounds coincide with the origin, one-half the squares of the radial distances or "errors" to the impact points of the bivariate circular normal distribution are distributed in an exponential fashion, and they can be ordered in magnitude if desired for analytical studies. In fact, if some of the rounds miss a circular target, they may be censored and the parameter of the exponential, and hence the circular normal distribution may be estimated by the use of Eq. 7-39. Moreover, if large errors of measurement are associated with the larger miss distances, as is often the case, or if the miss distances themselves perhaps follow a bimodal type of distribution due to the mixture of two different populations, the shots with the largest radial errors may be censored or truncated, and sample order statistic theory may be applied for estimation of the parameter. In fact, as we will see, the parameter σ^2 may be estimated from only one of the radial errors or some number of the inner ones without biasing results.

Coon (Ref. 28) has shown that for the ordered radial errors represented as

$$r_1 \leq r_2 \leq \cdots, r_i \leq \cdots \leq r_n \tag{7-50}$$

then the mean $E(r_i)$ and the second moment $E(r_i^2)$ about the origin of the ith order radial error are as follows:

$$E(r_i) = \sqrt{\frac{\pi}{2}}\sigma \sum_{k=0}^{i-1} \binom{n}{i-1-k}\binom{n-i+k}{k} \frac{(-1)^k}{(n-i+k+1)^{1/2}} \tag{7-51}$$

and

$$E(r_i^2) = 2\sigma^2 \sum_{k=0}^{i-1} \left(\frac{1}{n-i+k+1}\right) = 2\sigma^2 \sum_{k=1}^{i} \left(\frac{1}{n-i+k}\right). \tag{7-52}$$

The variance of the ith order radial distance therefore is given by

$$\sigma^2(r_i) = \mathrm{Var}(r_i) = E(r_i^2) - [E(r_i)]^2 \tag{7-53}$$

and the standard deviation of r_i is the square root of Eq. 7-53.

One would expect that the ordered radial deviations would be correlated so that a computation of the covariances in addition to the variances also would be of interest. In this connection Coon (Ref. 28) also has calculated the covariances, and we refer interested readers to her manuscript because such equations are rather complex.

For possible applications, we give in Table 7-7 the means and standard deviations of the radial order statistics from Coon's manuscript (Ref. 28) for samples of size through $n = 20$. We also reproduce her Table II as Table 7-8, which gives the variances and covariances of the ordered radii through the sample of size $n = 10$.

*Note that $\sigma^2 = \theta$ and is to be estimated. Moreover, t is exponential, but r is not.

TABLE 7-7

MEANS AND STANDARD DEVIATIONS OF THE ORDERED RADII IN A SAMPLE OF n FROM A CIRCULAR NORMAL DISTRIBUTION (Ref. 28)
(ALL ENTRIES ARE IN UNITS OF σ)

n	i	Mean	Std. Dev.	n	i	Mean	Std. Dev.
2	1	0.88623	0.46325	9	1	0.41777	0.21838
	2	1.62040	0.61180		2	0.64585	0.23473
					3	0.83495	0.24658
3	1	0.72360	0.37824		4	1.01199	0.25911
	2	1.21148	0.44608		5	1.18997	0.27428
	3	1.82486	0.58012		6	1.38008	0.29437
					7	1.59784	0.32378
4	1	0.62666	0.32757		8	1.87558	0.37433
	2	1.01443	0.37093		9	2.32579	0.49865
	3	1.40852	0.42747				
	4	1.96364	0.55747	10	1	0.39633	0.20817
					2	0.61072	0.22191
5	1	0.56050	0.29299		3	0.78637	0.23203
	2	0.89129	0.32497		4	0.94828	0.24228
	3	1.19915	0.35875		5	1.10756	0.25412
	4	1.54810	0.41236		6	1.27239	0.26889
	5	2.06753	0.54037		7	1.45187	0.28868
					8	1.66040	0.31780
6	1	0.51166	0.26746		9	1.92938	0.36801
	2	0.80468	0.29296		10	2.36983	0.49177
	3	1.06451	0.31647				
	4	1.33379	0.34785	11	1	0.37789	0.19753
	5	1.65526	0.40014		2	0.58078	0.21099
	6	2.14998	0.52686		3	0.74546	0.21982
					4	0.89546	0.22844
7	1	0.47371	0.24762		5	1.04071	0.23805
	2	0.73939	0.26897		6	1.18778	0.24954
	3	0.96789	0.28678		7	1.34289	0.26406
	4	1.19334	0.30819		8	1.51415	0.28363
	5	1.43913	0.33856		9	1.71525	0.31253
	6	1.74171	0.39007		10	1.97696	0.36245
	7	2.21803	0.51582		11	2.40912	0.48570
8	1	0.44311	0.23163	12	1	0.36180	0.18912
	2	0.68787	0.25009		2	0.55485	0.20155
	3	0.89396	0.26435		3	0.71038	0.20938
	4	1.09110	0.28028		4	0.85071	0.21679
	5	1.29559	0.30084		5	0.98497	0.22482
	6	1.52525	0.33062		6	1.11876	0.23410
	7	1.81386	0.38159		7	1.25680	0.24537
	8	2.27576	0.50657		8	1.40439	0.25972
					9	1.56903	0.27912
					10	1.76399	0.30783
					11	2.01956	0.35750
					12	2.44453	0.48029

(cont'd on next page)

TABLE 7-7 (cont'd)

n	*i*	Mean	Std. Dev.
13	1	0.34761	0.18170
	2	0.53213	0.19327
	3	0.67986	0.20031
	4	0.81214	0.20679
	5	0.93750	0.21365
	6	1.06092	0.22137
	7	1.18624	0.23046
	8	1.31728	0.24158
	9	1.45882	0.25579
	10	1.61800	0.27507
	11	1.80779	0.30361
	12	2.05806	0.35305
	13	2.47674	0.47543
14	1	0.33496	0.17509
	2	0.51199	0.18594
	3	0.65297	0.19234
	4	0.77843	0.19809
	5	0.89641	0.20404
	6	1.01146	0.21061
	7	1.12686	0.21816
	8	1.24561	0.22712
	9	1.37104	0.23813
	10	1.50759	0.25223
	11	1.66217	0.27138
	12	1.84750	0.29979
	13	2.09316	0.34903
	14	2.50624	0.47102
15	1	0.32360	0.16916
	2	0.49397	0.17939
	3	0.62096	0.18526
	4	0.74863	0.19041
	5	0.86037	0.19566
	6	0.96849	0.20134
	7	1.07591	0.20775
	8	1.18509	0.21518
	9	1.29856	0.22404
	10	1.41936	0.23497
	11	1.55171	0.24898
	12	1.70234	0.26803
	13	1.88379	0.29631
	14	2.12537	0.34536
	15	2.53345	0.46700
16	1	0.31333	0.16378
	2	0.47774	0.17348
	3	0.60760	0.17891
	4	0.72204	0.18358
	5	0.82841	0.18825
	6	0.93069	0.19325
	7	1.03151	0.19878
	8	1.13300	0.20508
	9	1.23718	0.21242
	10	1.34630	0.22119
	11	1.46320	0.23206
	12	1.59194	0.24600
	13	1.73914	0.26495
	14	1.91717	0.29311
	15	2.15511	0.34199
	16	2.55867	0.46330
17	1	0.30397	0.15889
	2	0.46301	0.16813
	3	0.58821	0.17317
	4	0.69811	0.17744
	5	0.79980	0.18165
	6	0.89706	0.18608
	7	0.99233	0.19093
	8	1.08748	0.19637
	9	1.18421	0.20258
	10	1.28427	0.20986
	11	1.38974	0.21849
	12	1.50328	0.22932
	13	1.62889	0.24319
	14	1.77305	0.26215
	15	1.94805	0.29015
	16	2.18272	0.33889
	17	2.58217	0.45989
18	1	0.29541	0.15442
	2	0.44957	0.16324
	3	0.57057	0.16796
	4	0.67642	0.17188
	5	0.77400	0.17570
	6	0.86690	0.17968
	7	0.95740	0.18397
	8	1.04722	0.18873
	9	1.13781	0.19410
	10	1.23062	0.20019
	11	1.32719	0.20753
	12	1.42952	0.21610
	13	1.54011	0.22724
	14	1.66303	0.24054
	15	1.80449	0.25953
	16	1.97677	0.28739
	17	2.20846	0.33603
	18	2.60415	0.45673

(cont'd on next page)

TABLE 7-7 (cont'd)

n	i	Mean	Std. Dev.	n	i	Mean	Std. Dev.
19	1	0.28753	0.15030	20	1	0.28025	0.14649
	2	0.43723	0.15876		2	0.42586	0.15462
	3	0.55443	0.16319		3	0.53959	0.15881
	4	0.65665	0.16682		4	0.63853	0.16219
	5	0.75056	0.17032		5	0.72915	0.16540
	6	0.83962	0.17391		6	0.81480	0.16868
	7	0.92599	0.17777		7	0.89754	0.17215
	8	1.01125	0.18194		8	0.97883	0.17594
	9	1.09667	0.18672		9	1.05986	0.18006
	10	1.18352	0.19187		10	1.14156	0.18544
	11	1.27302	0.19799		11	1.22544	0.18922
	12	1.36658	0.20541		12	1.31205	0.19577
	13	1.46615	0.21415		13	1.40280	0.20479
	14	1.57437	0.22402		14	1.49991	0.21375
	15	1.69462	0.23884		15	1.60620	0.22110
	16	1.83377	0.25705		16	1.72423	0.23626
	17	2.00358	0.28482		17	1.86122	0.25436
	18	2.23255	0.33344		18	2.02872	0.28245
	19	2.62480	0.45379		19	2.25521	0.33089
					20	2.64425	0.45104

TABLE 7-8

VARIANCES AND COVARIANCES OF THE ORDERED RADII IN A SAMPLE OF n FROM A CIRCULAR NORMAL DISTRIBUTION (Ref. 28)
(ALL ENTRIES ARE IN UNITS OF σ)

n	$i\backslash j$	1	2	3	4	5	6	7	8	9	10
2	1	0.2146	0.1348								
	2		0.3743								
3	1	0.1431	0.0957	0.0671							
	2		0.1990	0.1417							
	3			0.3365							
4	1	0.1073	0.0735	0.0551	0.0409						
	2		0.1376	0.1041	0.0777						
	3			0.1827	0.1379						
	4				0.3108						
5	1	0.0858	0.0596	0.0459	0.0363	0.0279					
	2		0.1056	0.0819	0.0651	0.0501					
	3			0.1287	0.1030	0.0796					
					0.1700	0.1327					
						0.2920					
6	1	0.0715	0.0500	0.0391	0.0318	0.0260	0.0204				
	2		0.0858	0.0674	0.0550	0.0450	0.0354				
	3			0.1002	0.0820	0.0674	0.0532				
	4				0.1210	0.0999	0.0791				
	5					0.1601	0.1277				
	6						0.2776				

(cont'd on next page)

TABLE 7-8 (cont'd)

n	$i\backslash j$	1	2	3	4	5	6	7	8	9	10
7	1	0.0613	0.0431	0.0340	0.0280	0.0235	0.0197	0.0157			
	2		0.0723	0.0573	0.0473	0.0398	0.0333	0.0266			
	3			0.0822	0.0682	0.0575	0.0482	0.0385			
	4				0.0950	0.0803	0.0674	0.0541			
	5					0.1146	0.0967	0.0778			
	6						0.1522	0.1234			
	7							0.2661			
8	1	0.0537	0.0379	0.0300	0.0250	0.0213	0.0182	0.0155	0.0125		
	2		0.0625	0.0498	0.0415	0.0354	0.0304	0.0258	0.0209		
	3			0.0699	0.0584	0.0499	0.0428	0.0364	0.0295		
	4				0.0786	0.0672	0.0578	0.0492	0.0399		
	5					0.0905	0.0780	0.0666	0.0541		
	6						0.1093	0.0936	0.0763		
	7							0.1456	0.1195		
	8								0.2566		
9	1	0.0477	0.0338	0.0269	0.0225	0.0194	0.0168	0.0146	0.0126	0.0103	
	2		0.0551	0.0440	0.0370	0.0318	0.0276	0.0240	0.0206	0.0169	
	3			0.0608	0.0511	0.0440	0.0383	0.0334	0.0287	0.0235	
	4				0.0671	0.0579	0.0505	0.0440	0.0378	0.0310	
	5					0.0752	0.0657	0.0573	0.0494	0.0405	
	6						0.0867	0.0758	0.0654	0.0537	
	7							0.1048	0.0907	0.0748	
	8								0.1401	0.1161	
	9									0.2487	
10	1	0.0429	0.0305	0.0243	0.0205	0.0177	0.0155	0.0137	0.0120	0.0104	0.0086
	2		0.0492	0.0395	0.0333	0.0288	0.0252	0.0223	0.0196	0.0170	0.0140
	3			0.0538	0.0455	0.0394	0.0346	0.0305	0.0268	0.0233	0.0192
	4				0.0587	0.0509	0.0447	0.0395	0.0348	0.0302	0.0249
	5					0.0646	0.0568	0.0502	0.0443	0.0385	0.0318
	6						0.0723	0.0640	0.0565	0.0491	0.0406
	7							0.0833	0.0737	0.0642	0.0532
	8								0.1010	0.0882	0.0733
	9									0.1354	0.1131
	10										0.2418

The covariance between r_i and r_j is

$$\mathrm{Cov}(r_i r_j) = E(r_i r_j) - E(r_i)E(r_j).$$

In 1952 Daniels (Ref. 29) published a paper on the probability distribution of the "covering circle" of a bivariate sample from a circular normal distribution. The covering circle is defined as the smallest circle in the xy plane that contains on it or inside it each and every sample point. In his paper Daniel (Ref. 29) points out the rather remarkable fact that the covering circle radius for a sample of n (rounds) from a circular normal distribution with mean (0,0) follows exactly the same distribution as the $(n-1)$st ordered radial error in a sample of n from the same circular normal distribution. Thus this provides a checkpoint with the work of Coon (Ref. 28) especially insofar as estimating the underlying sigma. We give an example concerning this point in the sequel.

For the circular normal distribution we see that the radial deviations or "errors" are of much importance in analyses of the precision and accuracy of firing weapons and, moreover, the circular error probable (CEP) is based on the equal sigma case for the mutually perpendicular directions. (CEP is defined as the radius of the circle about the shots, which includes half of the rounds.) On the other hand, for the unequal sigma case the analysis of radial errors and the CEP become much more difficult. Ref. 30 gives a thorough treatment of the various one- and two-dimensional measures of precision and accuracy of firing weapons, including standard errors in the two directions, the extreme horizontal and vertical dispersions (i.e., the univariate range), the mean horizontal and vertical dispersions, the radial standard deviation, the CEP, the mean radius, the extreme spread or bivariate range, the radius of the covering circle of Daniels (Ref. 29), and the "diagonal" of the shots. The unequal sigma cases are discussed as is the relative efficiency of the various measures of precision.

Perhaps the reader will now understand the importance of the sample order statistics—whether univariate, radial, etc.—to the general military requirement of analyzing the accuracy of fire of weapons of all types. Indeed, it is really the fact that one can truncate or censor some of the shots or radial deviations or can have them truncated for him by target misses(!) that becomes of much convenience and utility in the required statistical analyses. In fact, either all or some of the ordered radial errors can be used. In accordance with Eq. 7-50 and Table 7-7, only a single order statistic is really needed to estimate the sigma of the shots, or alternatively, some or all of the fixed, low number of the smaller radii can be used in accordance with Eq. 7-39. Of course, the precision or efficiency of estimation of sigma improves with and, in fact, depends on the number of sample order statistics actually used in the calculation. To illustrate this statement and as a case in point, refer to Table 7-7 for 10 rounds and only the smallest radial deviation. Thus for $n = 10$ and $i = 1$, we see that the mean of r_1 is about 0.396σ and the standard deviation is about 0.208σ. Hence by knowing r_1, the normal population sigma or σ may be estimated from $r_1/0.396 = 2.53r_1$, and the relative precision of this estimator is $0.208/0.396 = 0.53$. Had we used only the fourth smallest radius, the estimator of sigma would be $r_4/0.948 = 1.05r_4$, and the relative precision for the fourth smallest radial error improves to $0.242/0.948 = 0.26$, or $1/2$ that of r_1. If the largest radial error r_{10} is used to estimate sigma and it can be depended upon—i.e., is not a "wild" observation—then the precision of this estimator would be $0.492/2.370 = 0.21$, so that the gain is not so great at all now, and thus we see that some wild shots may be censored. Finally, had we used all 10 radial impacts and the estimator from Eq. 7-39, which becomes

$$\hat{\sigma} = (\hat{\sigma^2})^{1/2} = \left(\sum_{i=1}^{10} r_i^2/10\right)^{1/2}$$

then the relative precision of this estimator would be about 0.16.

We will further illustrate the use of the radial order statistics with Examples 7-8 and 7-9.

Example 7-8:

Given that the bullet impacts on a vertical target at 75 m follow a circular normal distribution and that all of the holes in the target from 10 shots can be inclosed in a circle of radius 6 in. Estimate the circular normal standard deviation of the population.

We will assume that the C of I of the rounds is centered on the origin point of the target. By using the result of Daniels (Ref. 29) that the covering circle radius for n shots follows the same probability distribution as the next to the largest radial error of the n impacts, we see from Table 7-7 for $n = 10$ and $i = 9$ that the mean value of the 9th ordered radial error is about 1.929σ. Therefore, our estimate of the population sigma is

$$\hat{\sigma} = 6/1.93 = 3.1 \text{ in.}$$

This result may be checked by noting in Daniels' paper (Ref. 29) or Table 7, p. 17, of Ref. 30 that the mean value of the radius of the covering circle for 10 shots is also given as 1.929σ

Example 7-9:

Find the chance that the largest radial deviation in a sample of eight shots on a target will exceed 3σ.

This is a somewhat more difficult problem because clearly the probability that the largest radius will exceed

3 sigmas is greater than the chance that any radial error at random will exceed this same limit. However, we note that Eq. 7-7 gives the chance that any number r of the ordered sample statistics will not exceed any stated value x. Therefore, we may take the cumulative probability of the distribution of interest up to 3 sigmas and then use Eq. 7-7 to obtain the desired chance. For the bivariate circular normal distribution, the cumulative probability to the 3σ point is

$$F(x) = F(3\sigma) = 1 - \exp[-x^2/(2\sigma^2)] = 1 - \exp(-9/2) = 0.98889.$$

Substituting this in Eq. 7-7 for $r = 8$ and $n = 8$, we obtain

$$I_{0.98889}(8,1) = 0.915$$

so that the correct probability that the largest of eight radial errors will exceed 3 sigmas is $1 - 0.915 = 0.085$.

Finally, for the treatment of radial errors, we should summarize the results for firing at vertical targets. In fact, there may seem to be some confusion because in the preceding account we have made use of both the radial errors (to the first power) and the squares of the radial errors. In practice, of course, it is generally easier to deal with the radial errors directly, i.e., their first powers, in making measurements on a target. We keep in mind, nevertheless, that our prime interest is in estimating the underlying, unknown sigma given in Eq. 7-45, but it is the square of sigma that relates directly to the chi-square distribution. In spite of this, we see that the underlying population sigma may be estimated from either the radial errors by using Coon's Table 7-7, or we can estimate the square of sigma by using the squares of the radial errors first and then by taking the square root. Moreover, the theory generalizes to any number of dimensions, say, p. Thus if $p = 2$, we are dealing with deviations from a C of I on a bivariate or plane target, whereas if $p = 3$, we analyze radial deviations from a C of I in three-space. Statistically, we consider that the radial deviation or "error" in p dimensions is represented by the quantity

$$r = (r_1^2 + r_2^2 + \cdots + r_p^2)^{1/2} \tag{7-54}$$

so that

$$r^2/\sigma^2 = \chi^2(p) \tag{7-55}$$

has the chi-square distribution with p df. Hence this means that the density function $f_p(r^2/\sigma^2)$ for p dimensions is

$$F(r^2/\sigma^2) = \{1/[\Gamma(p/2)2^{p/2}]\}\,(r^2/\sigma^2)^{p/2-1}\exp[-r^2/(2\sigma^2)]. \tag{7-56}$$

Now let us consider only the two-dimensional case, or $p = 2$, for target firings and a circular "target" of radius r_0. Here we may consider dealing with an actual circular target with C of I of rounds located at the target center, or we may want to analyze impacts on a target of any shape for which we arbitrarily truncate the use of impacts that are at a radial distance of more than r_0 units from the C of I of the rounds. The latter situation may be arrived at by simply finding a circle of radius r_0 about the impacts that includes as many of the impacts as possible, hoping, of course, that such a circle is "centered" for the impacts. In either case, it is seen that if all rounds are included in the equation to estimate the population variance (and hence none miss the "circular" target or are censored), the estimate of sigma for the complete sample of n rounds will be given by

$$\hat{\sigma} = [\sum_{i=1}^{n} r_i^2/(2n)]^{1/2} \tag{7-57}$$

On the other hand, if m of the rounds miss the circular target in n rounds fired or if we were to truncate m of the n rounds at a distance of r_0 units of the radial direction from the C of I, then sigma is estimated from the number $(n - m)$ of hits on the target, or

$$\hat{\sigma} = \{[\sum_{i=1}^{n-m} r_i^2 + (n-m)r_0^2]/[2(n-m)]\}^{1/2} \tag{7-58}$$

where only $(n - m)$ sample impacts are used in the calculation.

As a final point, it should be easily seen that the square of the quantity (Eq. 7-57) follows the chi-square distribution with $2n$ df, whereas the square of Eq. 7-58 is distributed as chi-square with $2(n - m)$ df. Accordingly, confidence bounds on either the unknown sigma squared or on sigma may therefore be determined.

In Ref. 31 Cohen gives some further treatment of the general problem in this area of radial impacts since he also treats the case in which the number of eliminated observations may be unknown. Here iteration techniques must be used to obtain the estimates of the unknown population sigma. Interested readers should consult Cohen's paper (Ref. 31). White (Ref. 32) also discusses radial errors.

As overall insight and some reflection, many readers will note that there clearly remains some needed research to be performed on the problem of analysis of radial errors. For example, it becomes very difficult to center the rounds on a desired target point or to guarantee that the C of I of the rounds is always at the particular aim point of interest. Moreover, the usual case is that the mean point of impact (C of I) has to be estimated from the observed sample values for the occasion. Thus we would certainly invite others to investigate such general problems much more deeply for the purpose of arriving at appropriate solutions. As a suggestion and in the present absence of an exact solution, perhaps an approximate chi-square technique (par. 4-4.5) might well be satisfactory on practical grounds. In this connection, an investigator might possibly consider also the Appendix, par. E, p. 25 of Ref. 30, for some ideas. Hopefully, our discussion on sample radial errors will stimulate others interested in the theory to attain results needed in applications.

7-10 PARAMETER ESTIMATION FROM TRUNCATED FIRINGS AT RECTANGULAR TARGETS

Although we have already said much about the estimation of the true unknown population mean and sigma for the important practical case in which samples are often truncated or censored for one reason or the other, there is considerably more to be said or discussed. We will therefore close out this particular subject with some further points of interest. In fact, we have really discussed only that part of the general problem that relates to the circular normal distribution and the use of radial distances. Fortunately, as we have already indicated, if we know there is truly a circular normal distribution, the radial errors may be used because even though the target may not be circular in shape, we may still truncate the sample firings at some given or fixed radial distance r_0 and estimate sigma according to Eq. 7-58. On the other hand, for the case of unequal, or suspected unequal, standard deviations in the x- and y-directions and also for the most usual case for which the targets are square or rectangular, some different methods of estimation have to be used.

Perhaps through a study or reading of this chapter so far, many readers will already have in mind some ideas concerning the estimation problem, previously discussed, for rectangular targets. For example, insofar as estimation of the normal population sigma is concerned, if it happens from target firings that the same number of rounds r miss the rectangular target on the left as on the right, or r miss below and r above, then simply the quasi-range of par. 7-3 could be used, especially to obtain a quick estimate of σ. However, one would not often be so fortunate as to have the same number missing on each side of the target, so that the sth largest and the rth smallest sample order statistics would be used, for example. Then again and better, for efficient estimation of the population mean and sigma, one would certainly consider the linear estimation techniques of pars. 7-5 and 7-6 and especially the type of problem illustrated in Example 7-4. Indeed, for independent or uncorrelated x- and y-directions separately, a rectangular target and the assumption of unequal sigmas in the two directions, one may obviously apply the estimation techniques of Example 7-4 in each direction alone and hence get good estimates of the mean x, the mean y, and the two sigmas in the x- and y-directions. Moreover, this may be done for quite unequal numbers of missing rounds above, below, to the left, and to the right of the target. We caution, of course, that these numbers should be known exactly; otherwise, some additional biases would be introduced.

Over the years of statistical investigations into the analyses of target firings, a number of techniques have been developed to handle this type of involved problem, which clearly requires and should be adapted to computer calculations. Cohen (Ref. 33) discusses the problem of restriction and selection in bivariate normal distributions and also the task of estimation in truncated bivariate normal distributions in another paper (Cohen, Ref. 34). Ref. 34 even considers the general bivariate normal case for which there exists a nonzero

population correlation coefficient. Hence these references of Cohen should provide some very worthwhile background material for the reader.

Our coverage of the problem of estimation for truncated bivariate normal samples against rectangular targets in this chapter is limited to a computer program (FORTRAN), which has been developed by Visnaw (Ref. 35) and used very satisfactorily as evidenced in a number of successful calculations for even a relatively large number of missing rounds. Visnaw's computer program is for the uncorrelated bivariate normal distribution, i.e., the case of independence of the x- and y-directions and for the estimation of the population means and the population variances of x and y. The details are covered in Visnaw's report (Ref. 35). He considers firings against a vertical target of width, say, 2A, and height 2B, on which is imposed a rectangular coordinate system with origin at the center of the rectangle. The horizontal distance x is then positive to the right of the origin, and the vertical direction y is taken as being positive in the upward direction. In an "accuracy" firing of rounds a total of N rounds are fired at the target with the result of possible impacts and misses as indicated on Fig. 7-4. Note that as a result of firing there are n_{22} rounds that actually impact the target and for which one might calculate biased ("deflated") values of the means in the two directions and the variances or standard deviations. Nevertheless, it is quite evident that due to missing rounds and the number of them, one would bias standard deviations toward the low side—and perhaps drastically—while the estimates of the true means of x and y would be shifted to the left, right, upward, or downward. Note also from Fig. 7-4 that nine rectangular regions or areas are considered with the numbers n_{ij} of rounds in each of the indicated areas for $i,\ j = 1,\ 2,\ 3$, and n_{11} is on the lower left part of the figure. In accordance with the notation of this chapter, we might say that for the x-direction there are r rounds missing on the left of the rectangular target, and the s largest values of x are truncated to the right of the target. In a like manner, we could say that there are g rounds missing below the target and h rounds above, so that

$$r = n_{11} + n_{12} + n_{13} = \text{Visnaw's } m_1 \tag{7-59}$$
$$s = n_{31} + n_{32} + n_{33} = \text{Visnaw's } m_3 \tag{7-60}$$
$$g = n_{11} + n_{21} + n_{23} = \text{Visnaw's } m_1' \tag{7-61}$$
$$h = n_{13} + n_{23} + n_{33} = \text{Visnaw's } m_3'. \tag{7-62}$$

In addition, Visnaw (Ref. 35) considers the number m'' of rounds, if needed, that indicates the number of rounds for which one is not able to sense a direction. In any event, there are a total of N rounds fired at the vertical target, which consists of the number n_{22} on the target and the stated categories in Eqs. 7-59 through 7-62 plus m'' if required.

If we refer to the impact coordinates of the actual hits on the target as (x_i, y_i), then for the totality of firings we are interested in estimating the population means μ_x and μ_y and the population standard deviations σ_x and σ_y. The details of the estimation procedures and the accompanying theory are covered by Visnaw in Ref. 35. His computer program in FORTRAN is listed here in Table 7-9. The statistical analysis of Visnaw (Ref. 35) involves ML estimation of the parameters, and he gives a numerical analysis in the form of an illustration, which we will reframe as our Example 7-10.

Example 7-10: (Based on numerical data of Visnaw in Ref. 35)

An accuracy test was conducted for a new recoilless rifle that consisted of firing 22 rounds at a vertical target 5 ft by 5 ft. Unfortunately, the target was placed too close to the gun so that there were only 14 impacts on the square target and the other eight rounds missed. Of the eight missing rounds, six missed to the left and are accounted for by

$$n_{11} = 3, \quad n_{12} = 1, \quad \text{and} \quad n_{13} = 2$$

but it was not possible to observe just where the remaining two rounds missed or passed the target. Irrespective of the eight missing rounds in 22 fired, it is required to obtain ML estimates of the true means and standard deviations of the overall population, assuming it is a bivariate normal distribution with possibly different sigmas in the two directions. The coordinates of the 14 impacts on the target surface are

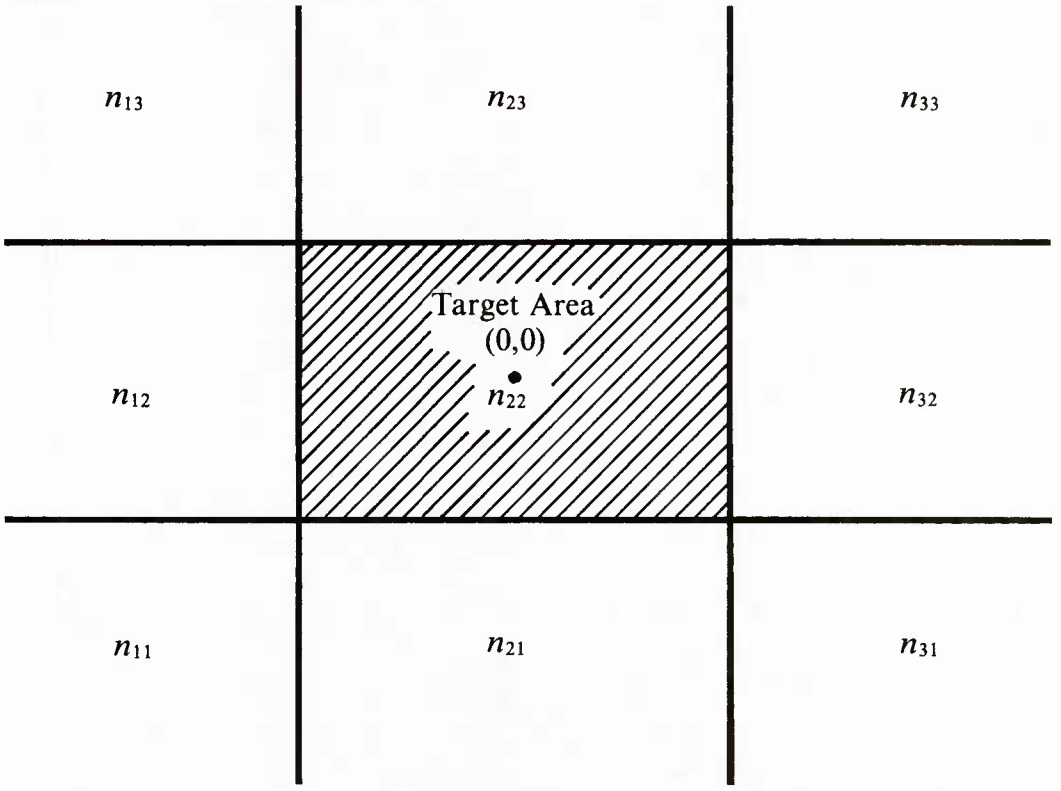


Figure 7-4. Schematic Diagram of Target and Areas of Missing Rounds

x	y	x	y
17.10	3.80	12.50	-6.10
23.70	28.60	6.50	2.70
12.80	-8.00	-1.60	15.00
29.50	18.80	4.20	18.10
11.20	15.60	1.90	10.00
-4.90	15.60	3.60	27.30
9.90	17.60	12.00	5.70.

If one were to use only the given 14 impact points to calculate the means and standard deviations in the x- and y-directions, the results would be:

	x	y
Mean:	9.886	11.764
Standard deviation:	9.050	10.636.

We would expect that the true mean point of impact would be to the left and below that for the estimate based on target hits since eight rounds missed in such a direction, more or less, and that the true standard deviations would be much underestimated. In fact, the calculations on a computer using the program of Table 7-9 gave the following estimates:

	x	y
Mean:	-7.610	6.977
Standard deviation:	29.035	25.954.

Thus there indeed is a shift in the direction suspected because of the probable location of missing rounds, and the standard deviations are underestimated by a factor of about 2.5 to 3. (Ref. 35 is available for interested users from the Analytical Branch, Materiel Test Directorate, US Army Test and Evaluation Command, Aberdeen Proving Ground, MD 21005.)

TABLE 7-9

COMPUTER PROGRAM FOR TRUNCATED BIVARIATE NORMAL TARGET FIRINGS (Ref. 35)

```
FORTRAN IV G LEVEL  21                MAIN                DATE = 79099          19/59/11

            C     ESTIMATION OF VERTICAL TARGET PARAMETERS
            C     AGNES M KODAT
            C     AUGUST 71
 C001               IMPLICIT REAL*8(A-H,O-Z)
 C002               DIMENSION X(100),Y(100)
 C003               DIMENSION CA(16),BA(4)
 C004         1C00  FORMAT(2F6.0)
 C005         1C05  FORMAT(I2)
 C006         1C10  FORMAT(I4)
 C007         1C15  FORMAT(F4.0)
 C008         1C20  FORMAT(I4)
 C009         1C25  FORMAT(4F8.0)
 C010         1C30  FORMAT(2F8.0)
 C011         1C40  FORMAT(4F10.0)
 C012         1C50  FORMAT('1'///T10,'ESTIMATION OF VERTICAL TARGET PARAMETERS'/T22,
                   1'(GENERAL CASE)'//)
 C013         1060   FORMAT(T10,'TARGET WIDTH=',F6.2/T10,'TARGET HEIGHT=',F6.2///)
 C014         1C70   FORMAT(T10,'NO. OF IMPACTS ON TARGET=',I4/)
 C015         1C80   FORMAT(T31,'BASED ON IMPACTS'/T24,'HORIZONTAL',T46,'VERTICAL'/
                   1T10,'MEAN',2F20.4/T6,'VARIANCE',2F20.4/T6,'STD.DEV.',2F20.4/////)
 C016         1C90   FORMAT(T10,'NO. OF SHOTS FIRED=',I4/)
 C017         2C00   FORMAT(T31,'ESTIMATED PARAMETERS'/T24,'HORIZONTAL',T46,'VERTICAL'/
                   1T10,'MEAN',2F20.4/T6,'VARIANCE',2F20.4/T6,'STD.DEV.',2F20.4//)
 C018         2C10   FORMAT('1'///T20,'IMPACT COORDINATES'//T17,'X(I)',T37,'Y(I)'//
                   1(2F20.2))
 C019         2C20   FORMAT(/////T20,'IT DID NOT CONVERGE'//)
 C020         2C30   FORMAT(///T5,'TARGET NO. =',F4.0)
 C021         2C40   FORMAT(///T17,'NO. OF ROUNDS MISSED THE TARGET'///T29,'MP3=',I4///
                   1T19,'N13=',I4,T29,'N23=',I4,T39,'N33=',I4///T29,'XXXXXXXX'/T29,
                   2'XXXXXXXX'/T10,'M1=',I4,T19,'N12=',I4,T29,'XXXXXXXX',T39,'N32=',I4
                   3,T49,'M3=',I4/T29,'XXXXXXXX'/T29,'XXXXXXXX'///T19,'N11=',I4,T29,
                   4'N21=',I4,T39,'N31=',I4///T29,'MP1=',I4///T29,'M''=',I4///)
 0022         2C50   FORMAT(/////T10,'STARTING VALUES'/T24,'HORIZONTAL',T46,'VERTICAL'/
                   1T10,'MEAN',2F20.4/T6,'VARIANCE',2F20.4/T6,'STD.DEV.',2F20.4////)
 C023               READ(5,1C00)A,B
 C024               A=A/2.DO
 C025               B=B/2.DO
 C026         1C0   READ(5,1010)N
 C027               IF(N-9998)110,250,250
 C028         110   READ(5,1C20)N11,N12,N13,N21,N23,N31,N32,N33,M1,M3,MP1,MP3,MDP
 C029               READ(5,1C05)ICODE
 C030               READ(5,1030)(X(I),Y(I),I=1,N)
 C031               JM=1
 C032               AN=N
 C033               LN=N+N11+N12+N13+N21+N23+N31+N32+N33+M1+M3+MP1+MP3+MDP
 C034               BN=LN-MDP
 C035               AW=2.DO*A
 C036               BH=2.DO*B
 C037               SUMX=0.DO
 C038               SUMY=0.DO
 C039               XSQ=0.DO
 C040               YSQ=0.DO
 C041               DO 120 I=1,N
 C042               SUMX=SUMX+X(I)
 C043               SUMY=SUMY+Y(I)
 C044               XSQ=XSQ+X(I)**2
 C045               YSQ=YSQ+Y(I)**2
 C046         120   CONTINUE
 C047               XM=SUMX/AN
 C048               YM=SUMY/AN
 C049               XVAR=(XSQ-AN*XM**2)/AN
 C050               YVAR=(YSQ-AN*YM**2)/AN
 C051               XSD=DSQRT(XVAR)
 C052               YSD=DSQRT(YVAR)
 C053               GO TO (125,126),ICODE
 C054         125   XX=SUMX+A*(N31+N32+N33+M3-N11-N12-N13-M1)
 C055               XXS=XSQ+(N31+N32+N33+M3+N11+N12+N13+M1)*A**2
 C056               YY=SUMY+B*(N13+N23+N33+MP3-N11-N21-N31-MP1)
```

(cont'd on next page)

TABLE 7-9 (cont'd)

```
C057              YYS=YSO+(N13+N23+N33+MP3+N11+N21+N31+MP1)*B**2
C058              XMM=XX/BN
C059              YMM=YY/BN
C060              XVV=(XXS-BN*XMM**2)/BN
C061              YVV=(YYS-BN*YMM**2)/BN
C062              GO TO 128
C063        126   READ(5,1025)XMM,YMM,XVV,YVV
C064        128   SDX=DSQRT(XVV)
C065              SDY=DSQRT(YVV)
C066              XMU=XMM
C067              YMU=YMM
C068              XSIG=SDX
C069              YSIG=SDY
C070        130   SX=C.DO
C071              SY=C.DO
C072              SXX=0.DC
C073              SYY=0.DC
C074              DO 140 I=1,N
C075              SX=SX+(X(I)-XMU)/XSIG
C076              SY=SY+(Y(I)-YMU)/YSIG
C077              SXX=SXX+((X(I)-XMU)/XSIG)**2
C078              SYY=SYY+((Y(I)-YMU)/YSIG)**2
C079        140   CONTINUE
C080              VX=XSIG**2
C081              VY=YSIG**2
C082              UL=(-A-XMU)/XSIG
C083              CALL PRNORM(UL,ZUL,A1)
C084              VL=(-B-YMU)/YSIG
C085              CALL PRNORM(VL,ZVL,B1)
C086              UU=(A-XMU)/XSIG
C087              CALL PRNORM(UU,ZUU,AA)
C088              A3=1.DO-AA
C089              A2=1.DC-A1-A3
C090              VU=(B-YMU)/YSIG
C091              CALL PRNORM(VU,ZVU,BB)
C092              B3=1.DC-BB
C093              B2=1.DO-B1-B3
C094              BN=N11+N12+N13+M1
C095              CN=N11+N21+N31+MP1
C096              DN=N31+N32+N33+M3
0097              EN=N13+N23+N33+MP3
C098              AB=1.DO-A2*B2
C099              FN=N21+N23-MDP*A2*B2/AB
C100              QN=N12+N32-MDP*A2*B2/AB
C101              FA=ZUL/A1
0102              FB=(ZUL-ZUU)/A2
0103              FC=ZUU/A3
C104              F=(-BN*FA+FN*FB+DN*FC+SX)/XSIG
0105              GA=ZVL/B1
0106              GB=(ZVL-ZVU)/B2
0107              GC=ZVU/B3
C108              G=(-CN*GA+QN*GB+EN*GC+SY)/YSIG
0109              HA=UL*ZUL/A1
C110              HB=(UL*ZUL-UU*ZUU)/A2
0111              HC=UU*ZUU/A3
0112              H=(-BN*HA+FN*HB+DN*HC+SXX-AN)/XSIG
0113              EA=VL*ZVL/B1
0114              EB=(VL*ZVL-VU*ZVU)/B2
0115              EC=VU*ZVU/B3
0116              E=(-CN*EA+QN*EB+EN*EC+SYY-AN)/YSIG
C117              GN=MDP*B2/A2
C118              GM=MDP*A2/B2
0119              FD=(ZUL-ZUU)/AB
0120              FE=(UL*ZUL-UU*ZUU)/AB
0121              FG=(ZUL*UL**2-ZUU*UU**2)/A2
0122              GD=(ZVL-ZVU)/AB
0123              GE=(VL*ZVL-VU*ZVU)/AB
0124              GG=(ZVL*VL**2-ZVU*VU**2)/B2
0125              HD=UL*ZUL/A2
C126              HE=UU*ZUU/A2
0127              ED=VL*ZVL/B2
```

(cont'd on next page)

TABLE 7-9 (cont'd)

```
0128                EE=VU*ZVU/B2
0129                FMX=(-BN*FA*(UL+FA)-GN*FD**2+FN*(HB-FB**2)+DN*FC*(UU-FC)-AN)/VX
0130                FMY=(-MDP*FD*GD)/XSIG/YSIG
0131                FSX=(-F*XSIG-BN*HA*(UL+FA)-GN*FD*FE+FN*(FG-FB*HB)+DN*HC*(UU-FC)-SX
                   1)/VX
0132                FSY=(-MDP*FD*GE)/XSIG/YSIG
0133                GMY=(-GN*GA*(VL+GA)-GM*GD**2+QN*(EB-GB**2)+EN*GC*(VU-GC)-AN)/VY
0134                GSX=(-MDP*GD*FE)/XSIG/YSIG
0135                GSY=(-G*YSIG-GN*EA*(VL+GA)-GM*GD*GE+QN*(GG-GB*EB)+EN*EC*(VU-GC)-SY
                   1)/VY
0136                HSX=(-H*XSIG-BN*HA*(HA+UL**2-1.D0)-GN*FE**2+FN*(HD*(UL**2-1.D0)-HE
                   1*(UL**2-1.D0)-HB**2)+DN*(HC*(UU**2-1.D0)-HC**2)-2.D0*SXX)/VX
0137                HSY=(-MDP*FE*GE)/XSIG/YSIG
0138                ESY=(-E*YSIG-GN*EA*(EA+VL**2-1.D0)-GM*GE**2+QN*(ED*(VL**2-1.D0)-EE
                   1*(VU**2-1.D0)-EB**2)+EN*(EC*(VU**2-1.D0)-EC**2)-2.D0*SYY)/VY
0139                CA(1)=FMX
0140                CA(2)=FMY
0141                CA(3)=FSX
0142                CA(4)=FSY
0143                CA(5)=FMY
0144                CA(6)=GMY
0145                CA(7)=GSX
0146                CA(8)=GSY
0147                CA(9)=FSX
0148                CA(10)=GSX
0149                CA(11)=HSX
0150                CA(12)=HSY
0151                CA(13)=FSY
0152                CA(14)=GSY
0153                CA(15)=HSY
0154                CA(16)=ESY
0155                BA(1)=-F
0156                BA(2)=-G
0157                BA(3)=-H
0158                BA(4)=-E
0159                CALL SIMQ(CA,BA,4,KS)
0160                IF(DABS(BA(1))-.001)150,150,200
0161          150   IF(DABS(BA(2))-.001)160,160,200
0162          160   IF(DABS(BA(3))-.001)170,170,200
0163          170   IF(DABS(BA(4))-.001)180,180,200
0164          200   XMU=XMU+BA(1)
0165                YMU=YMU+BA(2)
0166                XSIG=XSIG+BA(3)
0167                YSIG=YSIG+BA(4)
0168                JM=JM+1
0169                IF(JM-30)130,130,210
0170          180   WRITE(6,2010)(X(I),Y(I),I=1,N)
0171                WRITE(6,2040)MP3,N13,N23,N33,M1,N12,N32,M3,N11,N21,N31,MP1,MDP
0172                WRITE(6,1050)
0173                WRITE(6,1060)AW,BH
0174                WRITE(6,1070)N
0175                WRITE(6,1080)XM,YM,XVAR,YVAR,XSD,YSD
0176                WRITE(6,1090)LN
0177                WRITE(6,2000)XMU,YMU,VX,VY,XSIG,YSIG
0178                GO TO 100
0179          210   WRITE(6,2010)(X(I),Y(I),I=1,N)
0180                WRITE(6,2040)MP3,N13,N23,N33,M1,N12,N32,M3,N11,N21,N31,MP1,MDP
0181                WRITE(6,1050)
0182                WRITE(6,1060)AW,BH
0183                WRITE(6,1070)N
0184                WRITE(6,1080)XM,YM,XVAR,YVAR,XSD,YSD
0185                WRITE(6,2020)
0186                WRITE(6,2050)XMM,YMM,XVV,YVV,SDX,SDY
0187                GO TO 100
0188          250   STOP
0189                END

*OPTIONS IN EFFECT*   NOID,EBCDIC,SOURCE,NOLIST,NODECK,LOAD,NOMAP
*OPTIONS IN EFFECT*   NAME = MAIN     , LINECNT =       50
*STATISTICS*     SOURCE STATEMENTS =      189,PROGRAM SIZE =     8688
*STATISTICS*   NO DIAGNOSTICS GENERATED
```

(cont'd on next page)

TABLE 7-9 (cont'd)

```
C001            SUBROUTINE PRNORM(X,Z,P)
C002            IMPLICIT REAL*8(A-H,O-Z)
C003            T=DABS(X)
C004            IF(T-5.DO)10,10,5
C005        5   T=5.DO
C006       10   C=.398942280401
C007            D=.2316419
C008            B1=.31938153
C009            B2=-.356563782
C010            B3=1.781477937
C011            B4=-1.821255978
C012            B5=1.330274429
C013            Z=C*DEXP(-T*T/2.DO)
C014            V=1.DO/(1.DO+D*T)
C015            P=1.DO-Z*V*(B1+V*(B2+V*(B3+V*(B4+V*B5))))
C016            IF(X)15,20,20
C017       15   P=1.DO-P
C018       20   RETURN
C019            END
```

7-11 PARAMETER ESTIMATION FOR TRUNCATED POISSON SAMPLES WITH MISSING ZEROS

The use of order statistics and truncated sample theory go hand-in-hand as we have often illustrated in this chapter. Hence usually the need exists for joint studies of both statistical areas. Moreover, some form of order statistics and truncated sample theory is very often useful in dealing with practical problems relating to discontinuous distributions. A type of Army application we mentioned earlier for missing zero observations is certainly no exception in this connection. In fact, there exist many applications for which we have need of the Poisson or binomial distribution and for which, in practice, the number of zero observations is not observable. Some combat data, or other sampling experiments, are cases in point. Often it becomes desirable to study the results of combat data for purposes of inference, and the number of hits on targets, such as tanks that were knocked out in a battle, provides a useful and often typical illustration. After the battle is over one can survey the battlefield and try to gather analyzable data. However, the number of misses is not observable, and yet this figure would be important in establishing the total number of rounds fired in the battle, which in turn might be required to predict the chance that a fired round will result in a kill, or no kill, or this figure would be needed to predict logistical requirements of the total number of tank rounds needed in future battles, etc. Then again, there is the problem of estimating parameters of the assumed population in an unbiased manner even though the sample data were truncated. The problem for the Poisson distribution may be framed as indicated in the discussion that follows.

Since the chance of a hit overall against another tank on the battlefield is often small and the number of kills is not very large, we could safely assume that the number of hits per tank killed follows a Poisson distribution. Hence if we assume that the expected number of hits would be λ, the chance of exactly x hits would be given by

$$p(x) = \lambda^x \exp(-\lambda) / x! \tag{7-63}$$

whereas the chance of h or more hits would be determined from

$$P(h) = \sum_{x=h}^{\infty} \lambda^x \exp(-\lambda)/x!. \tag{7-64}$$

In 1959 Cohen (Ref. 36) investigated the estimation of the parameter λ by using Fisher's ML approach and found that the estimate $\hat{\lambda}$ of λ could determined from the equation

$$\hat{\lambda} \sum_{x=1}^{\infty} f_x / [1 - \exp(-\hat{\lambda})] = \sum_{x=1}^{\infty} x f_x \tag{7-65}$$

where f_x is the number of observed cases or frequencies for which exactly x hits occur, and the summation is stopped when the frequencies f_x are exhausted.

Note that only the frequencies for $x = 1, 2, 3$, etc., hits are included in the summation and that zero frequency, or number of rounds fired without any hits, is excluded, i.e., truncated from the estimation, Eq. 7-65. It should be clear that the solution of Eq. 7-65 for the unknown λ is rather easily obtainable on a pocket calculator by cut-and-try methods.

Cohen (Ref. 36) points out that regardless of the value of the true unknown parameter λ, the asymptotic variance of the estimate satisfies the equation

$$\lambda/n \leq \sigma^2(\hat{\lambda}) \leq 2\lambda/n \tag{7-66}$$

where n is the total number of observations or the total frequency included in Eq. 7-65, but it does not include the unknown number f_0 for the zero class. In Ref. 36 Cohen did not address the problem of the estimation of f_0, an important parameter nevertheless. The estimation of the zero class frequency was undertaken by Cohen in 1960 (Ref. 37) and later by Dahiya and Gross (Ref. 38) in 1973 for the truncated Poisson distribution by using ML estimation techniques, and their procedure gives an estimate, first, of the total number of observations including the zero frequency. Thus the estimate of f_0 for the zero class frequency is obtained from the equation

$$N = f_0 + f, \qquad f = \sum_{x=1}^{\infty} f_x \tag{7-67}$$

and the estimate of N, the grand total, from

$$\hat{N} = \sum_{x=1}^{\infty} x f_x / \hat{\lambda}. \tag{7-68}$$

Thus and regardless of the fact that the frequency for the zero class is not observable in many applications, we nevertheless can obtain efficient estimates of the parameter λ from Eq. 7-65 and also the proper estimate of the zero class frequency f_0 from Eqs. 7-68 and 7-67.

These ML estimators of the Poisson parameter λ and the zero class frequency f_0 turn out to be quite satisfactory although recently some further investigation has been done on the estimation problem by Blumenthal, Dahiya, and Gross (Ref. 39). They develop the ML and modified ML estimators further and investigate their asymptotic properties theoretically in some detail, as well as making use of Monte Carlo experiments to judge comparisons. Also an example is given in Ref. 39 concerning the use of the new estimators.

In Example 7-11 we will illustrate the application of the truncated Poisson distribution in relation to a combat survey to collect data for further inferences concerning tank engagements.

Example 7-11:

A major battle broke out in Western Europe between Blue's First Army and Red's Fifth Army with a series of tank battles over a wide landscape. For this conflict Blue decided to attach many additional tanks to its force since it believed that tanks would be the key striking arm that would win the battle, as Blue did. Nevertheless, Blue decided to conduct an analysis of Red's capability to engage and destroy Blue tanks and, in particular, to estimate the expected number of Blue tank kills due to Red in a typical battle and decide on just how many rounds total Red may have fired at Blue tanks in such an engagement. After the battle subsided, a Blue military operations research team surveyed the battlefield and made a count of the number of Blue tanks killed and of the number of hits on each killed tank. The latter figures were taken, for past experience had shown that a single armor-piercing projectile hit on a tank would normally result in a kill.

The number of Blue tanks with $x = 1$ or more hits that were knocked out of the battle and the frequency f_x of hits per killed tank are given in Table 7-10.

To obtain an efficient estimator of the the expected number of Blue tanks that may be killed per Red antitank round fired, we note using the second equation of Eq. 7-67 that

$$f = 65 + 22 + 3 + 1 = 91.$$

TABLE 7-10

BLUE TANKS WITH ONE OR MORE HITS AND OBSERVED FREQUENCY

x Number of hits per Blue tank	f_x Number of tanks with x hits
1	65
2	22
3	3
4	1
5	0

Then, calculating the RHS of Eq. 7-65, we see that

$$\sum_{1}^{5} x f_x = 65(1) + 22(2) + 3(3) + 1(4) = 122.$$

Thus by cut-and-try methods with Eq. 7-65, we obtain the estimate

$$\hat{\lambda} = 0.618$$

or in other words, the tank battle was so intense and at such close range that Red's potent antitank guns would be expected to kill 0.62 Blue tanks per round!

With reference to the number of rounds fired by Red that missed we use Eq. 7-68 and obtain $N = 122/0.618 = 197.4$, so that

$$f_0 = 197 - 91 = 106 \text{ rounds.}$$

For the information and further enlightenment of the reader (or even the ubiquitousness of order statistics!), the observed data in Table 7-10 actually were taken from a study and classical example of Bortkiewicz (Ref. 40), which describes the number of deaths from kicks of horses in the Prussian Army during the period 1875 to 1894, except that the number of exposures for which no deaths occurred was reported to be 109 as compared to the 106 estimated! Furthermore, if one now uses the complete sample including the frequency 109 of the zero class, the estimate of the Poisson parameter results in

$$\hat{\lambda} = 122/(109 + 91) = 0.61$$

an expected number very close to that predicted from the truncated sample, which was $\hat{\lambda} = 0.618$!

7-12 SUMMARY

In this chapter we have attempted to bring together some of the more basic and useful tools concerning the analysis of sample order statistics that the Army statistician will have occasion to apply in his work. These include—but are not limited to—the sample range, the distribution of the largest and smallest sample values, the quasi-ranges, the expected values of sample order statistics, linear estimation of population means and standard deviations from truncated samples, the statistics of extremes, relations to the outlier testing problem, tolerance intervals of general distributions, the relation of order statistics to reliability analyses, radial order statistics in target accuracy analyses, the estimation of parameters from truncated target firings, and the truncated Poisson distribution. Several examples were given to illustrate the applicable theory.

REFERENCES

1. H. Leon Harter, *Order Statistics and Their Use in Testing and Estimation*, Volume 1, *Tests Based on Range and Studentized Range of Samples from a Normal Population*, Aerospace Research Laboratories, Office of Aerospace Research, United States Air Force, 1969 (For sale by the US Government Printing Office, Washington, DC 20402).
2. H. Leon Harter, *Order Statistics and Their Use in Testing and Estimation*, Volume 2, *Estimates Based on Order Statistics of Samples from Various Populations*, Aerospace Research Laboratories, Office of Aerospace Research, United States Air Force, 1969 (For sale by the US Government Printing Office, Washington, DC 20402).
3. H. Leon Harter, *A Chronological Annotated Bibliography on Order Statistics*, Volume 1: *Pre-1950*, Air Force Wright Aeronautical Laboratories, Air Force Systems Command, Wright-Patterson Air Force Base, OH, August 1977.
4. A. E. Sarhan and B. G. Greenberg, *Contributions to Order Statistics*, John Wiley and Sons, New York, NY, 1962.
5. Herbert A. David, *Order Statistics*, John Wiley and Sons, New York, NY, 1970.
6. L. H. C. Tippett, "On the Extreme Individuals and the Range of Samples Taken from a Normal Population", Biometrika **17**, 364-87 (1925).
7. E. S. Pearson and H. O. Hartley, Editors, *Biometrika Tables for Statisticians, Vol. I*, Cambridge University Press, Cambridge, England, 1954.
8. William C. Guenther, "An Easy Method of Obtaining Percentage Points of Order Statistics", Technometrics **19**, 319-21 (August 1977).
9. J. H. Cadwell, "The Distribution of Quasi-Ranges in Samples from a Normal Population", The Annals of Mathematical Statistics **24**, 603-13 (1953).
10. F. E. Grubbs and C. L. Weaver, "The Best Unbiased Estimate of Population Standard Deviation Based on Group Ranges", Journal of the American Statistical Association **42**, 224-41 (June 1947).
11. E. H. Lloyd, "Least Squares Estimation of Location and Scale Parameters Using Order Statistics", Biometrika **39**, 88-95 (1952).
12. D. B. Duncan, "Multiple Range and Multiple *F* Tests", Biometrics **11**, 1-42 (1955).
13. D. Teichroew, "Tables of Expected Values of Order Statistics and Products of Order Statistics for Samples of Size Twenty and Less from the Normal Population", The Annals of Mathematical Statistics **27**, 410-26 (June 1956).
14. Gunnar Blom, *Statistical Estimates and Transformed Beta-Variates*, John Wiley and Sons, Inc., New York, NY, 1958.
15. A. E. Sarhan and B. G. Greenberg, "Estimation of Location and Scale Parameters by Order Statistics from Singly and Doubly Censored Samples", The Annals of Mathematical Statistics **27**, 427-51 (June 1956).
16. H. A. David, "Some Uses of Order Statistics", Proceedings of the Twentieth Conference on the Design of Experiments in Army Research Development and Testing, Part 2, US Army Research Office Report No. 75-2, 833-9, June 1975.
17. E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, NY, 1958.
18. N. R. Mann, R. E. Schafer, and N. D. Singpurwalla, *Methods for Statistical Analysis of Reliability & Life Data*, John Wiley and Sons, New York, NY, 1973.
19. Samuel S. Wilks, "Determination of Sample Size for Setting Tolerance Limits", The Annals of Mathematical Statistics **12**, 91-6 (1941).
20. AMCP 706-109, Engineering Design Handbook, *Tables of the Cumulative Binomial Probabilities*, June 1972.
21. R. B. Murphy, "Non-Parametric Tolerance Limits", The Annals of Mathematical Statistics **19**, 581-9 (1948).
22. Paul N. Somerville, "Tables for Obtaining Non-Parametric Tolerance Limits", The Annals of Mathematical Statistics **29**, 599-601 (1958).

23. Albert H. Bowker, "Tolerance Limits for Normal Distributions", Chapter 2, *Techniques of Statistical Analysis*, by the Statistical Research Group, Columbia University, McGraw-Hill and Company, Inc., New York, NY, 95-110 (1947).
24. Albert H. Bowker, "Computations of Factors for Tolerance Limits on a Normal Distribution When the Sample is Large", The Annals of Mathematical Statistics **17**, 238-40 (1946).
25. S. K. Mitra, "Tables for Tolerance Limits for a Normal Population Based on Sample Mean and Range or Mean Range", Journal of the American Statistical Association **52**, 88-94 (1957).
26. DARCOM-P 706-101, Engineering Design Handbook, *Army Weapon Systems Analysis, Part One,* November 1977.
27. B. Epstein and M. Sobel, "Life Testing", Journal of the American Statistical Association **48**, 486-503 (1953).
28. Helen J. Coon, "The Sample Order Statistics of a Radial Distribution", Manuscript. Available from the US Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD, 1967.
29. H. E. Daniels, "The Covering Circle of a Sample from a Circular Normal Distribution", Biometrika **39**, 137-43 (1952).
30. Frank E. Grubbs, *Statistical Measures of Accuracy for Riflemen and Missile Engineers*, Edwards Brothers, Ann Arbor, MI, 1964 (Copies available from the author).
31. A. C. Cohen, Jr., "Maximum Likelihood Estimation of the Dispersion Parameter of a Chi-Distributed Radial Error from Truncated and Censored Samples With Applications to Target Analysis", Journal of the American Statistical Association **50**, 1122-35 (1955).
32. Robert G. White, *Distribution and Moments of Radial Error*, NASA Technical Memorandum NASA TM X-64692, Marshall Space Flight Center, AL, 15 August 1975.
33. A. Clifford Cohen, Jr., "Restriction and Selection in Samples from Bivariate Normal Distributions", Journal of the American Statistical Association **50**, 884-93 (1955).
34. A. Clifford Cohen, Jr., *Estimation in Truncated Bivariate Normal Distributions*, Technical Report No. 2, Army Research Office Contract No. DA-01-009-ORD-288 with the University of Georgia, Athens, GA, 30 June 1953.
35. Vernon V. Visnaw, *A Method of Estimating Center of Impact and Dispersion for Groups of Projectiles With Partially Missing Impact Coordinates*, Analytical Branch Report 71-AS-106, Material Test Directorate, USATECOM, Aberdeen Proving Ground, MD, 6 October 1971.
36. A. Clifford Cohen, Jr., *Estimating the Poisson Parameter from Truncated Samples With Missing Zero Observations*, Technical Report No. 15 (Revised), Army Research Office Contract No. DA-01-009-ORD-463 with the University of Georgia, Athens, GA, 3 August 1959.
37. A. Clifford Cohen, Jr., "Estimation in Truncated Poisson Distributions When Zeros and Some Ones are Missing", Journal of the American Statistical Association **55**, 139-43 (June 1960).
38. Ram C. Dahiya and Alan J. Gross, "Estimating the Zero Class from a Truncated Poisson Sample", Journal of the American Statistical Association **68**, 731-3 (1973).
39. S. Blumenthal, R. C. Dahiya, and A. J. Gross, "Estimating the Complete Sample Size from an Incomplete Poisson Sample", Journal of the American Statistical Association **73**, 182-7 (March 1978).
40. L. von Bortkiewicz, *Das Gesetz der Kleinen Zahlen*, Teubner, Leipzig, Germany, 1898.

CHAPTER 8

DETERMINATION OF SAMPLE SIZES

The problem of sample size determination represents—for both the present and future—one of the most important requirement applications for the practicing Army statistician or analyst. We therefore introduce and discuss the analytical problem area in sufficient detail to present a good introduction and to encourage additional research of the general subject.

It is often true that the determination of sample sizes on the basis of statistical grounds alone is insufficient and that the engineering or physical aspects must frequently be brought to bear. This is especially the case for some very high-reliability requirements and also for the investigation of critical, but low-chance, types of materiel defects. Nevertheless, statistical considerations do indeed solve, in a very satisfactory manner, many of the problems of sample size determination faced by the Army.

In this chapter we present the various methods of determining sample sizes for the common statistical tests of significance and introduce the estimation of sample sizes for designs of experiments. The subject is approached either by requiring a high level of confidence that an important or stated difference will be detected or, better still and secondly, by controlling errors of rejecting the statistical null hypothesis when it is true or accepting a false null hypothesis when an alternative is true.

Many examples are given to illustrate the theory.

8-0 LIST OF SYMBOLS

a = hypothesized or stated value of μ

b = particular calculated value or constant (see Eq. 8-62)

c = allowable number of failures in a binomial sampling plan

d = stated difference of interest to detect when calculating the sample size

$E(\hat{\theta})$ = mean value of $\hat{\theta}$

F = Snedecor-Fisher variance ratio or statistic

F_i = a theoretical frequency (relative to f_i)

$F_{1-\alpha}$ = upper α probability level of F

F_β = lower β probability level of F

f_i = observed or preliminary frequency

H_0 = null hypothesis that is tested for acceptance

H_1 = alternative hypothesis (to H_0) of special interest

k = ratio of sigmas or number of classes in a contingency table

m = number of normal populations sampled

n = number of observations in the sample

n_a = number of observations on which χ_a^2 is based

n_1 = size of the "first" sample, so designated

n_2 = size of the "second" sample, so designated

P_i = expected proportion

p = true unknown proportion in a binomial population

p_i = ith preliminary proportion

p_0 = null hypothesis H_0 value of the binomial parameter, usually representing the "acceptable" fraction

p_1 = alternative hypothesis H_1 value of the binomial parameter, representing the "unacceptable" fraction

$\tilde{p}$ = computed value of p that determines the critical region
$\hat{p}$ = preliminary observed proportion
$\hat{p}$ = estimate of p from the sample
$\hat{p_i}$ = occurrence ratio for estimating the binomial parameter p for a sample from the ith population $= x_i/n$
q = stated percentage of θ
r = number of failures
s = sample standard deviation based on $(n-1)$ degrees of freedom
s_a^2 = "new" sample variance based on the divisor $(n-3)$ instead of $(n-1)$
s_1^2 = sample variance based on (n_1-1) degrees of freedom for sample number 1
s_2^2 = sample variance based on (n_2-1) degrees of freedom for sample number 2
t = Student's t statistic
t = random variable, e.g., for an exponential distribution
$t_{1-\alpha}$ = upper α probability level of Student's t
Var( ) = variance of quantity in ( )
x_i = observed number of occurrences in a sample of n from ith binomial population
x_{ij} = ith item drawn at random from the jth normal population
$\bar{x}$ = sample mean (based on sample of size n)
$\bar{x}_1$ = sample mean of the first sample
$\bar{x}_2$ = sample mean of the second sample
$\bar{x}_{\cdot j}$ = sample average for the jth population
$\bar{x}_{\cdot\cdot}$ = grand (sample) average for all mn observations when m normal populations are sampled with n from each
$Z = (1/2)\ln F$ = Fisher's Z statistic
z = standard normal deviate, i.e., random variable from $N(0,1)$
z_α = upper α probability level of the standard normal deviate z. (It could better be designated as $z_{1-\alpha}$, so we call it $+z_\alpha$.)
z_β = standard normal deviate associated with the Type II error β. (usually taken as $+z_\beta = z_{1-\beta}$)
$\tilde{z}$ = value of the standard normal deviate z, which determines the boundary of the "critical" region
α = chance of rejecting the null hypothesis H_0 if true (also Type I error)
β = chance of accepting H_0 if the alternative H_1 were true (Thus, β is the Type II error.)
δ = ratio (of mean lifetime parameters) as in Eq. 8-72
η = ratio as in Eq. 8-74
θ = hypothesized fraction of sigma (see par. 8-7)
θ = mean lifetime parameter of the exponential distribution as in Eq. 8-66
θ_0 = hypothesized value of θ under H_0
θ_1 = hypothesized value of θ under H_1
$\hat{\theta}$ = angle in radians, determined from the arc sine transformation
$\hat{\theta}$ = best estimate of the mean lifetime parameter
λ = ratio of specified mean failure times θ_0/θ_1 as in Eq. 8-71
λ = ratio of two unknown population standard deviations
λ = expected number of occurrences for a Poisson distribution

λ = ratio of the desirable to the undesirable standard deviations, equal to the ratio of two chi-squares as in Eq. 8-28
λ_0 = expected number under the null hypothesis H_0
λ_1 = expected number under the alternative hypothesis H_1
λ_1 = expected number of occurrences for the first Poisson population
λ_2 = expected number of occurrences for the second Poisson population
λ' = particular value of λ (see Eq. 8-59)
$\tilde{\lambda}$ = calculated value of lambda giving the boundary of the critical region
μ = true unknown mean of (usually) a normal population
μ_0 = specified or stated value of the population mean μ (for H_0)
μ_1 = value of the true mean if an alternative hypothesis H_1 is true
μ_1 = specified value of μ under H_1
σ = true unknown standard deviation of a population
σ_0 = hypothesized value of sigma under H_0
σ_1 = hypothesized value of sigma under H_1
σ_1 = true unknown standard deviation of the first population
σ_2 = true unknown standard deviation of the second population
χ_a^2 = "available" value of chi-square (e.g., from past data)
χ_d^2 = desired or projected significant value of chi-square
$\chi_\gamma^2(\nu)$ = lower γ probability level of chi-square with ν degrees of freedom (df)

8-1 INTRODUCTION

One of the questions most frequently asked of the statistician is, "What sample size should I use?". A simple question, but one that usually has a very complex answer! In fact, all kinds of "qualifications" are really required even to begin to answer this question, and perhaps the universal answer could be "You get what you pay for." Nevertheless, we will explore this general question in some detail since the practical man must have some kind of suitable answer, and he wants some assurance that his experiment will not have been for naught. Clearly, just any or a small sample size will often fail to detect a real difference between treatments, and obviously, there is no need to use a large sample size to attain a definitive conclusion that could have been attained with only a fraction of the effort expended. Stated another way, one would like to have high confidence or high assurance that his conclusions from an experiment will be valid and that they could be used for prediction purposes in future, similar, or even more general situations.

Sample size is necessarily tied in with the population variance or standard deviation. Thus if one is sampling a population of interest and desires to estimate the true mean with "precision", the observed sample mean, or even an individual observation, could be "close" to the parent mean if the individual observations in the population are "tight" or close together. On the other hand, if they are spread out or there is much "randomness", enough sampling has to be done so that the observed sample mean will exhibit suitable stability. Since the variance of a sample mean is equal to the population variance divided by the sample size, i.e., σ^2/n, it follows that—since in practically all cases nothing can be done to decrease or reduce to zero the population sigma—one must use the proper sample size to control or deal with the sampling variation or randomness.

Another factor that must be taken into consideration is the size of the difference one would like to detect or how close he would like to get to the true value. Thus one could have very high confidence with any sample size, but the width of the confidence interval within which one states that the population parameter lies depends on the sample size and, in fact, very markedly so. If the underlying sigma is relatively small, perhaps a "practical" sample size would be sufficient to detect a relatively small difference; otherwise, increasing sample sizes would be needed to make valid judgments. In estimating the population mean, for example, one might decide to perform enough tests to be able to state with 99% assurance that the observed sample mean is within a preset, small interval or distance of the unknown population mean. Therefore, he controls the width of the

confidence interval to some desired value, but he must take both the sigma and the sample size into proper consideration. At this point, we should emphasize that the population sigma is not known, nor may one have much detailed knowledge about its actual size, so that the confidence statement or interval may depend on fuzzy thoughts about this "nuisance" parameter, which invariably is included in the analysis! It is for such reasons that the statistician seeks to find and use confidence intervals that are void of nuisance parameters if at all possible. The reader will recall, for example, that if one is sampling a normal population and uses the Student's *t* test to make inferences about the size of the population mean, only sample data — i.e., the sample mean, sample standard deviation, and sample size *n*—are involved with the result that a nonnuisance parameter confidence interval can be placed about the unknown normal population mean. However, this very desirable, or "complete", or "sufficient", analytical statement is not valid for so many other needed applications. In any event, we see so far that in trying to establish principles for the determination of sample size, we must deal with precision or variance and also concern ourselves with the size of a difference to be considered, which hopefully relates to practical significance.

In the determination of sample size, one could simply focus on controlling the precision of the estimator to a given (small) size, i.e., the value of the variance of the estimator, for example, or he could determine the sample size such that the estimate of the parameter will be within a stated percentage of the true value. Such procedures to determine the sample size almost invariably run into the problem of needing to know much about the size or value of the population parameter, or especially the true variance. If one desires to estimate the fraction of defective articles in a binomial population, the precision of the estimator depends on the true fraction or percentage, which causes somewhat of a stumbling block to arise. For another case, if one is sampling a normal universe to estimate the true sigma, he could use the chi-square distribution, and with the aid of the sample variance, he could determine the sample size, or more exactly the number of degrees of freedom (df) in the sample estimate of variance, either to control the width of a confidence interval or to guarantee some high level of assurance that his estimator will be within some stated percentage of the true population parameter. Thus controlling precision might be easy if the sigma were known with sufficient accuracy, at least for many sampling situations involving various populations.

As a very useful and statistically sound procedure of estimating sample sizes, one of the more elegant and important practical ways is to determine the sample size so that errors of judgment are controlled, provided that sample sizes so determined are "reasonable" and "practicable". With this approach we are dealing with the power or the "operating characteristic" of the test. "Power" ordinarily deals with the probability that the statistical test of significance will very infrequently reject the null hypothesis when it is true, but it will, on the other hand, detect and hence reject the null hypothesis when it is false by some given amount. The test that rejects the null hypothesis when false with the greater frequency is the more powerful test. We will explore this formulation of the problem more in the sequel, and it is based on statistical theory developed by Jerzy Neyman and Egon S. Pearson. In connection with basing determination of the sample size on the power of the test, we are all aware that there are risks associated with sampling and, furthermore, that one cannot make a 100% positive judgment unless the whole universe is sampled. Obviously, one cannot afford to sample the entire population due to costs, or in the case of destructive tests such extensive sampling would be prohibitive because it would leave no useful items.

In any introductory account or discussion of sample size determination, we should mention the possibility that the selection of sample sizes based on only statistical procedures might, at least in some or even many cases, lead to what some managers refer to as "impractical" amounts of sampling or turn out to be too costly otherwise. Thus we enter a domain of thought that might prohibit the exercise of appropriate "statistical power". It is in such cases that "one gets what he pays for". We realize that restrictions on sampling must be invoked by management at times, but hopefully in situations of this kind suitable tests can be conducted that ordinarily will be sufficiently meaningful for the practical problem at hand. In fact, sometimes it might be possible to conduct some kind of sequential procedure to save on costs or sample size, and the Army statistician must be constantly aware of this possibility.

With this background concerning the problem of determination of the most appropriate sample sizes in test procedures, we will now describe a number of cases and methods. Wherever possible, we will give some informative examples to highlight the principles involved. First, we will discuss the problem of sampling binomial type populations and then proceed to the sampling of continuous distributions and the various

significance testing techniques that are commonly used in this area of application. As might be expected, practically all of the sample size determination problems in one way or another lead to the existence of normality or asymptotic normality for many sampling procedures. This applies, for example, to many binomial sampling problems. Therefore, it would be wise to record initially the technique that will be useful in such cases. We will now illustrate this point so the reader will acquire immediately some useful orientation of the methodology; then we will proceed to apply the procedure of asymptotic normality to the determination of binomial sample sizes and will later check on the accuracy of the normality assumption for particular cases.

Finally, and before proceeding to the various technical or statistical details involved, we mention in a preliminary way some of the more useful references the Army analyst might keep in mind for his various applications. Again, as for other statistical endeavors, there exists a large volume of accounts into the investigation of sample sizes that is scattered widely in the statistical literature. It is our problem to cite and expose some of the more important tools for the Army analyst.

As a source of some preliminary and already published recommendations for the Army analyst on sample size determination, the Engineering Design Handbooks (Refs. 1 and 2) contain some very valuable curves for sample size choice (much of which is based on or originated from Ref. 3 and, in particular, the operating characteristic curves of that paper). Thus Refs. 1, 2, and 3 all contain continuing contributions—which the Army analyst will long have use for—to the sample size determination problem for the more common statistical tests of significance. In this chapter we will repeat only those power curves or attainments considered necessary to make this chapter as complete as need be without requiring the joint use of any other material.

The paper of Chand (Ref. 4) also contains many very useful equations for the determination of sample size, especially from the hypothesis testing point of view, i.e., the control of Type I and Type II errors.

The American Society for Testing and Materials (ASTM) "practice" (Ref. 5) gives an illustration of some sample size determination problems, perhaps more closely allied with actual practice in industry, and its approach might be said to determine sample size based on significance tests or to detect a difference of some size of interest with high confidence, say 95%.

Although Ref. 6 had as its primary aim the design of single sampling inspection plans to control errors of judgment in Army testing, the content of that paper really addresses the problem of sampling a binomial population to test the hypothesis that the true proportion of defectives, or the proportion of successes, is some stated desirable fraction as compared to an undesirable fraction of occurrences. A very similar problem is addressed by Clark in Ref. 7 through use of the incomplete beta function of Karl Pearson.

For some "popular" or very practical approaches and for background education for those in applied fields, Hahn's papers (Refs. 8 and 9) contain some very informative points on sample size. A confidence level alone is not sufficient!

It might be said that Refs. 1-9 contain some of the basic procedures for sample size determination in the day-to-day task of the Army analyst who must deal often with tests of significance or who might be required to suggest a design of experiment that likely will produce clear-cut conclusions. On the other hand, there is a very large area of application for sample size selection, which applies to the more complex statistical experiments, and references to the pertinent literature for such applications will be covered as required in the sequel.

8-2 THE ROLE OF THE NORMAL DISTRIBUTION IN SAMPLE SIZE DETERMINATION

As previously stated, we will discuss the choice of sample size for several different approaches, but one of the usual procedures is to seek a normally distributed statistic or one that is approximately or asymptotically normal and to make calculations of the power of the test in detecting shifts in the normal population mean. Thus this procedure sets a relatively low risk, such as 0.05 or 0.01, for rejecting the null hypothesis when it is true and then requires calculations of the probabilities of rejecting or accepting the untrue null hypothesis when an alternative hypothesis, indicated by a shift in level, is actually true. The more quickly or the more frequently the null hypothesis is rejected when a shift in level occurs, the more powerful the test is, and this depends on the particular statistical test used, the variance of the statistic, and perhaps most of all on the sample size. We may easily illustrate this analytically by dealing with normally distributed variates and the use of the sample mean when the population sigma is assumed known. Here, we start with the standardized normal variate z given by

$$z = (\bar{x} - \mu)/(\sigma/\sqrt{n})] \tag{8-1}$$

where

$\bar{x}$ = sample mean
σ = the standard deviation of an individual observation
μ = true unknown mean
n = sample size

and the individual observations x_i are from $N(\mu,\sigma)$.
Next, we take the null hypothesis H_0 to be

$$H_0 : \mu = \mu_0$$

or that is, the true unknown mean of the normal population is a stated value μ_0. We then take the probability level of the test to be α, and the observed value of z for the assumed population $\mu = \mu_0$ substituted in Eq. 8-1 is compared with (in absolute value) the size of the upper normal percentage point $+z_\alpha$; the null hypothesis is rejected when that probability level is exceeded. If in fact we were to draw a sample of size n from the normal population with mean μ_0, then our chance of accepting μ_0—the true state of affairs—is clearly

$$Pr[-z_\alpha < z < z_\alpha] = (1/\sqrt{2\pi})\int_{-z_\alpha}^{+z_\alpha} \exp(-t^2/2)dt = 1 - 2\alpha.* \tag{8-2}$$

The Type I error, or chance of rejecting the null hypothesis when it is true, is therefore 2α (for this formulation of a two-sided test), since this is the probability that a random normal deviate will fall outside the significance level points.

Now let us suppose that H_0 is false and that actually an alternative hypothesis H_1 is true, i.e., the real state of affairs is that H_1 holds, where

$$H_1 : \mu = \mu_1 > \mu_0$$

where

μ_1 = value of true mean if an alternative hypothesis H_1 is true.

Such a situation could have resulted, for example, from a shift in the population mean or perhaps from the fact that we are ignorant of the true mean of the normal parent we are sampling. In any event, our calculation of the quantity z will now be in error because we would use μ_0 instead of the correct true mean μ_1. We can nevertheless calculate the true chance of accepting the null hypothesis when it is false by entering the normal tables with the correctly assumed mean μ_1, which again "centers" the normal population sampled. Therefore, the chance of accepting the false H_0 when H_1 is true can be correctly calculated by using new limits on the integral or, in other words, from the probability statement:

$$\beta = Pr[-z_\alpha + \sqrt{n}|\mu_1 - \mu_0|/\sigma < z < +z_\alpha + \sqrt{n}|\mu_1 - \mu_0|/\sigma]** \tag{8-3}$$

where z is the correctly centered and standardized normal deviate, and the normal tables would be entered with the new limits in Eq. 8-3. We have used the quantity β to designate the new probability of accepting the null hypothesis H_0 when H_1 is true. Note in particular that $\beta \neq (1 - 2\alpha)$ unless H_0 is true, i.e., the true mean of the normal population sampled is $\mu = \mu_0 = \mu_1$ also.

Examination of Eq. 8-3 shows that the distance between its end points in the probability statement is still $2z_\alpha$ as in Eq. 8-2, although we see also that when $\mu_1 \neq \mu_0$ the larger the sample size is the more "magnification" or the larger is the shift in the population mean, so to speak. This means that, whereas the Type I error—when

*For brevity we often use $z_{1-\alpha} = +z_\alpha$.

**Usually and when no confusion should arise, we will take z_α to be the upper positive α probability level of $N(0,1)$ although strictly $z_\alpha = -z_{1-\alpha}$.

H_0 is true—equals 2α, or that the chance of accepting a true null hypothesis H_0 is $(1-2\alpha)$, the probability β of accepting the incorrect H_0 when H_1 is actually true decreases rapidly with an increasing sample size n. Thus many readers will regard the hypothesis testing approach as being rather "negative" because we first set up a "straw man", or null hypothesis, which is very infrequently rejected when true. However, when a significant level of the test is attained using the observational data, we decide that such a result is "so rare" or unexpected that we must reject the null hypothesis and accept the alternative hypothesis as the correct state of affairs, so to speak. Nevertheless, an advantage of such an approach is that we can set a low chance of a Type I error—rejecting the null hypothesis when it is true—and furthermore control the Type II error—accepting the null hypothesis when false and an alternative is very likely—with the proper choice of sample size. Indeed, this approach certainly appears to be a very sound one especially if the sample sizes are not "impracticable".

For the test relating to the assumption of a normally distributed statistic z, we see that either the two-sided test of Eq. 8-2 or a one-tail test is used to set the significance level at the desired value; Eq. 8-3 is then used to find the operating characteristic (OC) curve (one minus power) of the two-tailed test. The OC curve is a graph of the chance of accepting the null hypothesis, as in Eq. 8-3, against all possible values of the normal population mean μ; thus the OC exhibits the "power" of the test, including sample size effect. OC curves, or one minus them, which give the power curves, are now widely used as aids in the determination of sample size.

In summary, we see from Eqs. 8-2 and 8-3 that an important relation exists between the Type I and Type II errors, the standard deviation σ, the true difference between possible population means, or $\mu_1 - \mu_0$, and the sample size n. This relationship, solved for the sample size n, is given by

$$n = \frac{\sigma^2(z_\alpha + z_\beta)^2}{(\mu_1 - \mu_0)^2} \tag{8-4}$$

where

z_β = standard normal deviate associated with the Type II error β.

Thus the sample size to guarantee a Type I error of only α* and a Type II error of only β* for accepting the null hypothesis H_0 when actually H_1 is the true situation—is the product of the variance of the normal population sampled and the square of the sum of the two upper percentage points of the normal distribution representing the Type I and Type II errors divided by the square of the difference in population means to be detected. A fairly easy way to prove this is by using a one-sided test involving the upper α probability level for the test of whether the true unknown normal population mean has shifted to an unacceptably large value. For this particular case, the rejection or "critical" region is given by an observed value of z from Eq. 8-1, which exceeds the percentage point z_α of the normal distribution when H_0 is true. However, if H_1 is true and the true mean has shifted to a higher level, the chance of rejecting the null hypothesis when false must be calculated from an observed z exceeding the negative of the left-hand side (LHS) of the inequality in the probability statement of Eq. 8-3, which is really z_β. Equating these two and solving for n gives Eq. 8-4.

As it turns out, the sample size n could be determined from a much more general problem. In fact, we not only could have let the mean level shift, but also have a change in the variance. That is, if the variance under the null hypothesis is σ_0^2 but the correct state of affairs is H_1 for which the variance is σ_1^2, the sample size should be determined from

$$n = \frac{(z_\alpha \sigma_0 + z_\beta \sigma_1)^2}{(\mu_1 - \mu_0)^2} \text{ **} \tag{8-5}$$

Of course, we need to know not only the sample size but also have at hand a general equation for the critical region, and this is based on a value of z given by the quantity

$$z > \tilde{z} = \frac{z_\alpha \mu_1 \sigma_0 + z_\beta \mu_0 \sigma_1}{z_\alpha \sigma_0 + z_\beta \sigma_1} \text{ **} \tag{8-6}$$

*α and β are both small, i.e., $\leq$ approximately 0.10.

**See, for example, the paper of Chand (Ref. 4).

where

$\tilde{z}$ = value of the standard normal deviate z, which determines the boundary of the "critical" region

which should be exceeded for the one-sided test when trying to guard against a higher population mean than expected. Hence we see that if we can reduce the sample size problem to that of using an approximately distributed normal variate, the equations connecting the key parameters are rather simple. In fact, we will show that Eqs. 8-4, 8-5, and 8-6 are very useful even for a discrete variable, such as a binomial one. Otherwise, we will apply these normal approximations since they will be of value in the determination of sample sizes for continuous random variables wherever appropriate.

Finally, for cases in which the required sample size equations are not so simple or it is otherwise convenient, graphs may be constructed from which the needed sample sizes may be read with facility. This, in fact, is just what has been done for the treatment of the sample size problem insofar as that covered by Refs. 1 and 2. Also OC curves are generally used in Ref. 3 and many other pertinent publications instead of equations.

8-3 SAMPLE SIZES AND CRITERIA FOR BINOMIAL- AND POISSON-TYPE DATA

8-3.1 SAMPLING A SINGLE BINOMIAL OR POISSON POPULATION

When one draws a single random sample of size n from a binomial universe to test a hypothesis concerning the value of the unknown parameter p, representing the true proportion of successes, or failures, etc., he will often be interested in the size of the sample to be taken and the acceptance criteria. By this, we mean that, for example, a sample of size $n = 20$ will be drawn, and we conclude that the proportion of defectives in the population is no more than, say, 5% if no more than one defective is found in the sample. This, in fact, is the sampling plan of the test or the acceptance sampling plan. We will illustrate this binomial-type sample size problem by using (1) a "significant difference" equation, such as that given in Ref. 5, then (2) the control of Type I and Type II errors approach with the normal approximation of par. 8-2, and finally (3) the direct or exact solution, especially for some comparisons.

Some investigators have advocated the use of the significance level of a test to establish the needed sample size. In fact, this approach is rather widely used, as indicated in the "standard recommended practice" of Ref. 5. For this approach of determining sample size, it is evident that the Type I error is taken into consideration, but no mention is made of the Type II error, which, of course, is greatly influenced by the sample size and its effect on the power of the test. Apparently, this approach was conceived and used from the standpoint of determining sample size for the purpose of estimating a key parameter only and not to control errors of judgment based on Type I and Type II errors. For this approach, and the use of binomial-type data, one takes the quantity z

$$z = (\hat{p} - p)/[p(1-p)/n]^{1/2} \tag{8-7}$$

where

$\hat{p}$ = sample success (or failure) ratio as an estimate of p
p = true unknown binomial parameter
n = sample size.

as being normally distributed.
With the quantity z, Eq. 8-2 is used in the form

$$Pr[-z_\alpha < z < z_\alpha] = Pr[|z| < z_\alpha]^* = Pr\{|(\hat{p} - p)/[p(1-p)/n]^{1/2}| < z_\alpha\}. \tag{8-8}$$

Finally, the two sides of the inequality in the very last probability statement of Eq. 8-8 are equated and the sample size n is solved for, i.e.,

*Since this is a two-sided test, it is at the 2α level. To guard against a high p, use the upper level only.

$$n = z_\alpha^2 p(1-p)/(\hat{p}-p)^2. \tag{8-9}$$

Clearly, one has to have knowledge of, or estimate, p which may be taken as the null hypothesis value p_0 and also needs to specify just how close the sample estimate must be to the population value p, i.e., the quantity $(\hat{p}-p)$.

Example 8-1 will illustrate just how Eq. 8-9 might be used for sample size determination.

Example 8-1:

In acceptance sampling procedures for Army tests of mechanical time fuzes, the practice is to take a random sample from each lot, assemble them to high explosive (HE) projectiles, and fire them from a gun for both the estimation of the percentage of duds and their timing precision and accuracy. A dud rate of not over about 1% was considered acceptable, and most manufacturers were apparently meeting this requirement. On the assumption that it was desired to estimate the dud rate within 1% from the sample fired, how large a sample should be taken to do so? To guard against a high dud rate, use the upper 5% level.

For the stated problem, it is easily seen that the sample size would be based on Eq. 8-9 and is

$$n = (1.645)^2(0.01)\,(0.99)\,/\,(0.01)^2 = 268.$$

Actually, it is believed that this is a very large sample size for the particular problem stated, and one should question whether the cost of the test is too high! We will, therefore, reframe the question in Example 8-2.

As a point of interest, we note in passing that Eq. (3) of Ref. 5 indicates the use of 3 instead of the 1.645 we have used; the 3 is for the upper 0.3% level of the normal distribution. Had we used 3, the required sample size would have been 891—a prohibitive value indeed!

Now let us reframe the problem requirements in terms of the use of Type I and Type II error protections and see what this turns out to be in practice.

Example 8-2:

Suppose in Example 8-1 we had, in addition to the data given, simply said that we certainly could not tolerate 10% duds in mechanical time fuzes and would like to reject lots of such fuzes at least 90% of the time.

This new formulation of the problem clearly calls for further and more detailed practical and statistical insight into the use of mechanical time fuzes. Furthermore, we have now set an "acceptable" and an "unacceptable" level, and we control the errors that are to be allowed in the sampling plan. For this particular formulation, we see that Eq. 8-5 is required, and we have

$$p_0 = 0.01,\ p_1 = 0.10,\ \alpha = 0.05,\ \beta = 0.10,\ z_\alpha = 1.645, \text{ and } z_\beta = 1.282.$$

Hence by applying Eq. 8-5, one calculates that the sample size is determined from

$$n = \left[\frac{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p_1(1-p_1)}}{p_0 - p_1}\right]^2 \tag{8-10}$$

where

p_0 = null hypothesis H_0 value of the binomial parameter p, which represents the "acceptable" fraction
p_1 = alternate hypothesis H_1 value of the binomial parameter p, which represents the "unacceptable" fraction

and for our particular hypothesized problem, we find $n = 37$, a very acceptable value. In this connection, one might argue that the $p_1 = 0.10$ has been set too high. If, for example, we were to use $p_1 = 0.05$ as the greatest unacceptable value, then we would determine that the sample size should be 123, which is perhaps a more reasonable value than $n = 268$.

To complete the sampling inspection plan, we substitute, in Eq. 8-6 to obtain $p=\tilde{p}$, which is determined from the equation

$$\tilde{p} = \frac{z_\alpha p_1\sqrt{p_0(1-p_0)} + z_\beta p_0\sqrt{p_1(1-p_1)}}{z_\alpha\sqrt{p_0(1-p_0)} + z_\beta\sqrt{p_1(1-p_1)}} \tag{8-11}$$

where

$\tilde{p}$ = computed value of p, which determines the critical region.

For the data of this example, we calculate $\tilde{p}=0.037$, which, if multiplied by the sample size of $n=37$, gives an acceptance number of 1.36. It will be found that if we use the acceptance sampling plan $c=1$ (a whole acceptance number) and $n=37$, then when the true fraction defective of the sampled lot is 0.01, the chance of the lot passing is 0.947 (i.e., only a single or zero defectives are found in 37 items inspected), and when the lot of fuzes has 10% duds, the chance of it passing the sampling plan is only 0.104—the values 0.947 and 0.104 are determined from the OC curve. Thus these probabilities are as close to the desired risk values as can be obtained with discrete binomial-type data.

In contrast to this asymptotic normal approach, Guenther (Ref. 10) has shown that both the sample size n and the acceptance number c may be obtained simultaneously and with great accuracy by using the chi-square statistic. It should be very clear that the chi-square approach makes much sense because the Poisson distribution is the same as the chi-square distribution for the case of an even number of df, and the Poisson distribution is an excellent approximation to the binomial distribution for either small or large p and large values of n. Actually, Guenther (Ref. 10) apparently has shown that the chi-square procedure may be very accurate in determining n and c for values of $p \leq 0.50$, which means that all values of p from zero to one will be covered by working with the parameter $(1-p)$ instead of p when necessary. Guenther's chi-square technique is based on the inequality

$$(1/2)[(1/p_1 - 0.5)\,\chi^2_{1-\beta}(2c+2) + c] \leq n \leq (1/2)[(1/p_0 - 0.5)\,\chi^2_{\alpha}(2c+2) + c] \tag{8-12}$$

where χ^2_γ is the (lower) γ probability level of chi-square with ν df.

To use Eq. 8-12, one takes the lower α level and the upper β level of the percentage points of the chi-square distribution with the given p_0 and p_1, and then by trial with different c's, or $(2c+2)$ df, finally finds the interval that contains at least one integer—the value of n. For example, applying Eq. 8-12 to the data of Example 8-2 for $p_1 = 0.10$, we may first try $c=0$, for which we get

$$(1/2)[(4.61)\,(9.5)] = 21.9 \leq n \leq (1/2)\,[(0.103)\,(99.5)] = 5.12$$

which obviously does not work. But next trying $c = 1$, we find the inequality

$$(1/2)\,[(7.78)\,(9.5) + 1] = 37.46 \leq n \leq (1/2)\,[(0.711)\,(99.5) + 1] = 35.87,$$

which shows that the n's are close, but c should perhaps be just greater than 1. If we use $c=2$, we find for Eq. 8-12 that we obtain

$$51.5 \leq n \leq 82.6$$

indicating that we have gone too far above the proper value of c.

We thus conclude as before that the correct plan is $c=1$, $n=37$.

As a further comment on the two different methods for determining the sample size, we see that the "significant difference" approach requires "a good estimate" of the true unknown p and leaves matters unresolved. The control of Type I and Type II errors approach, on the other hand, requires some very clear thought as to just what true p is really acceptable and what is unacceptable. This means better engineering or

physical insight into the problem at hand. Nevertheless, this latter approach is very desirable it would seem and certainly gives a "concrete" answer. In such a hypothesis-testing situation, one does have to come to a decision concerning the problem requirements!

Since we have used a normal approximation to determine the sample size and also the acceptance number (approximately), it should be asked whether such an approach is good enough. In this connection, Grubbs (Ref. 6) and later Clark (Ref. 7) set up this binomial sampling plan on a more "exact" basis by using percentage points of the binomial distribution or the incomplete beta function. In Ref. 6 there are two tables; Table I gives values of the true $p = p_0$ for which the sample size n and acceptance number c are such that the probability of accepting the lot is 0.95 (the upper 5% point), and Table II gives values of $p = p_1$ with accompanying n and c such that the chance of passing the plan is only 0.10, or the lower 10% point of the binomial distribution. Thus one may enter Table I with $p = p_0$ (= 0.01 in our case) and search for the n and c using also Table II for which $p = p_1$ (= 0.10 for our problem) which are the same. When, for the same n and c the two conditions are satisfied, the sampling inspection plan is determined. For Example 8-2 it will be found that the best or closest plan is indeed $n = 37$ and $c = 1$. Therefore, the normal approximation gives the "exact" answer, and hence there would seem to be no need to construct such extensive tables as those in Refs. 6 and 7 for this particular purpose, unless both p's are too "small" for the normal approximation.

In case p_0 and p_1 both do not exceed approximately 0.10, the Poisson approximation applies very adequately. We will demonstrate this by using Table III of Ref. 6, which is reproduced here as Table 8-1.

TABLE 8-1

95% AND 10% PROBABILITY LEVELS FOR THE POISSON DISTRIBUTION

Acceptance Number c	Values of np_0 for 95% Point	Values of np_1 for 10% Point
0	0.0513	2.303
1	0.3554	3.890
2	0.8177	5.332
3	1.366	6.681
4	1.970	7.994
5	2.613	9.275
6	3.285	10.53
7	3.981	11.77
8	4.695	12.99
9	5.425	14.21
10	6.169	15.41
11	6.924	16.60
12	7.690	17.78
13	8.464	18.96
14	9.246	20.13
15	10.04	21.29

Reprinted with permission from "On Designing Single Sampling Inspection Plans" by Frank E. Grubbs, Annals of Mathematical Statistics XX, No. 2 (June 1949). Copyright© by Institute of Mathematical Statistics.

To use Table 8-1, one merely has to divide the 95% Poisson points by the acceptable level p_0 for $c = 0, 1, 2$, etc., and the 10% points by the unacceptable p_1; the results are the sample sizes. Wherever the sample sizes "cross", as they will for some value of the acceptance number c, the sample size to use with that particular c value is determined. For illustration, if we use the data given for Example 8-2, then we compute for $c = 0$, 1, and 2:

$c = 0$	$c = 1$	$c = 2$	(only through $c = 1$ needed)
0.0513/0.01 = 5.1,	0.3554/0.01 = 35.5,	0.8177/0.01 = 81.8	for n's
2.303/0.10 = 23.0,	3.890/0.10 = 38.9,	5.332/0.10 = 53.3	for n's.

These simple calculations show that the sample sizes calculated cross at approximately $c = 1$; moreover, we should probably use $n = (35.5 + 38.9)/2 = 37.2$ or $n = 37$ as before. It is seen, therefore, that even though the acceptable p is small, i.e., 0.01, and the unacceptable $p = 0.10$ is not so small, the normal approximation is still quite good. In the sequel we will give another approximation that is quite good, especially for small values of p or high reliability; it is the arc sine approximation. Before turning to that, however, we should point out that Table 8-1 is limited in scope because it is for a Type I error of 0.05 and a Type II error of 0.10 only. In this connection, one may take a more complete table of the percentage points of the chi-square distribution, such as that of the *Biometrika Tables for Statisticians* (Ref. 11), enter it for even numbers of df, and divide each value of chi-square for a given percentage point by two in order to extend Table 8-1 to all desired probability levels or various degrees of protection. Thus this approach could be very useful either for an "exact" type of calculation for small p's or as a check on approximation equations. Guenther's chi-square (Ref. 10) is also very accurate and useful.

Another good approximation to be used in connection with small or high p's (high reliabilities) is the arc sine transformation often applied in the analysis of variance (ANOVA) techniques. It is well-known that for small values of p the angular approximation

$$\hat{\theta} = 2\mathrm{Sin}^{-1}\sqrt{\hat{p}}, \text{ rad} \tag{8-13}$$

is very nearly normally distributed with mean value

$$E(\hat{\theta}) = 2\mathrm{Sin}^{-1}\sqrt{p} \tag{8-14}$$

and variance

$$\mathrm{Var}(\hat{\theta}) \approx 1/n. \tag{8-15}$$

Of some particular note for the arc sine transformation of small percentages is the fact that the mean value is of the same form as the transformation itself and that a very desirable feature is that the variance depends only on the sample size and not on the nuisance parameter p at all. The sample size to control Type I and Type II errors to sizes α and β, respectively, for the case of sampling a single binomial (Poisson) population is rather widely known to be

$$n = (1/4)\left(\frac{z_\alpha + z_\beta}{\mathrm{Sin}^{-1}\sqrt{p_1} - \mathrm{Sin}^{-1}\sqrt{p_0}}\right)^2 * \tag{8-16}$$

and the critical region is determined from the quantity

$$\tilde{p} = 2\left(\frac{z_\alpha \mathrm{Sin}^{-1}\sqrt{p_1} + z_\beta \mathrm{Sin}^{-1}\sqrt{p_0}}{z_\alpha + z_\beta}\right)^* \tag{8-17}$$

or, that is, the acceptance number is taken as the lowest integer in $[n\tilde{p}]$.

Although both Eqs. 8-10 and 8-16 are asymptotically normal statistics, it should not be expected that they give exactly the same sample size n although the values they do give are sufficiently close together. We will give an example of possible uses of Eq. 8-16 that applies to the problem of sample size determination for the investigation of prematures, safety, or high reliability types of items or components under test.

Example 8-3:

Suppose the premature rate of only 1/100,000 for artillery projectiles is desired, and a rate of 1/1000 would be considered unacceptable. (Of course, we would like the premature rate to be zero, but it is not possible to always manufacture projectiles so that no prematures would ever occur.) Determine how large a sample

*The angles are in radians.

should be tested to guarantee that a lot of projectiles with the objectionable rate of 1/1000 would be rejected with 95% assurance and that a lot sampled with only 1/100,000 prematures would be accepted 95% of the time.

As a preliminary statement, we remark that it is not easy for the manager or engineer to set such rates, especially for prematures as those given. We merely are illustrating just what the sample size problem might be on a statistical basis only in order to discover the resulting economic implications.

The reader may verify, using Eq. 8-16, that the required sample size is about 3339. The allowable number of prematures is zero using Eq. 8-17 and checking the Type I error at the acceptable rate of 1/100,000. Hence in an actual test of the items, one would stop the firing as soon as a single premature occurred and reject the null hypothesis at that stage of testing since there would be no point in firing all 3339 rounds.

In view of this, it would seem that we have arrived at such a large sample size, at least for some items, that a statistically determined sample size cannot be afforded. Moreover, since the 1/100,000 and 1/1000 are "relatively far apart" and we have set risks at a "sizable" level, i.e., 1 in 20, for any tighter conditions the required sample size would be astronomical indeed. Thus often it may be the case that the testing of very large sample sizes becomes prohibitive, and in fact, nothing might be learned in such testing because the basic problem may be one of design. Therefore, once a critical defect such as a premature is observed and the frequency appears too great, one must delve into the item design problem to try to correct the engineering fault. In this connection, a combination of engineering and statistics will often result in designing test programs for the purpose of examining each possible cause of a premature that the design judgment might indicate. An interesting account of the investigation into the possible engineering causes of prematures for artillery projectiles is given by Simon (Ref. 12) in his discussion of the relation of engineering to very high reliability. In fact, since our Example 8-3 concerns prematures, i.e., safety problems, we will by contrast also include an example on high reliability insofar as sample sizes are relevant (Example 8-4).

Example 8-4:

Suppose we desire a reliability of 0.9999 for proper launch of the Gemini vehicle, and the National Aeronautics and Space Administration (NASA) expert judgment arrives at the conclusion that a failure rate of 1 in 1000 could not be tolerated. Before we put a man in the capsule, how many items would have to be tested to assure that this high degree of reliability is guaranteed?

For illustrative purposes, we might again start with a risk of, say, 5% of rejecting the "acceptable" design and a risk of 5% of accepting the undesired reliability of 0.999. By using Eqs. 8-16 and 8-17 and checking the Type I and Type II errors by computation, one finds that the acceptance sampling plan should be $c = 2$, $n = 5784$. Had we reduced the errors of classification from 5% down to 1%, the sample size would have to be 11,570 ($c = 4$)! Therefore, just how has NASA solved this type of problem? The answer almost has to be by sound technological considerations, excellent engineering, quality control, the use of redundant components, good simulation experiments, extensive testing of components, perhaps accelerated life-type tests or tests of increased severity, and elaborate checkout methods. In summary, high reliability and safety should begin with the actual design of a system and follow through the development, fabrication, and the testing of system parts. Statistical techniques, including the design of experiments and sample size determination, are an aid to management.

In addition to our account so far of sample size determination for safety- and high reliability-type problems, there is a sequential method of testing that might result in some savings of effort. This is based on stopping the test at the event of a single critical defect or failure, and in addition, it indicates just what the lower confidence bound on reliability would be at a point of stopping for which no failures or critical defects such as prematures have occurred. That is, one continues to sample with only the occurrence of "successes" and decides to stop at some point because of already having expended a large number of items in the test. The method to which we refer is covered on pp. 21-9 of the *Army Weapon Systems Analysis* Handbook (Ref. 13). Table 8-2 is repeated from Ref. 13 for the reader's use.

By reference to Table 8-2 we see that if in a test one attains 50 successes and no failures, it can be stated that the lower 95% confidence bound on the reliability of the item tested is 94.0% Had we achieved 400 successes with no failures in the 400 trials, then the lower 95% confidence bound on reliability would be 99.3%, etc. Note how slowly the bound rises for increasing numbers of tests as shown on Fig. 8-1. For example, in going from a sample size of 1000 to 2000, the increase in the lower confidence bound is only 15 in 10,000. Perhaps this adds

TABLE 8-2

LOWER 95% CONFIDENCE BOUNDS ON RELIABILITY BASED ON ZERO FAILURES IN n TRIALS

Number of Tests n	Lower 95% Confidence Bound on Reliability
50	0.940
100	0.970
200	0.985
300	0.990
400	0.993
500	0.994
1000	0.997
2000	0.9985
3000	0.9990
4000	0.9993
5000	0.9994
29957	0.9999

some insight into the formidable problem of guaranteeing very high reliability. Also it "drives one back to the need for genius in design problems"!

Although we have given sample size equations for drawing a single random sample from a binomial population with a small percentage of occurrences, we should nevertheless deal with the sampling of a Poisson population. Generally, the parameter of the Poisson population, which we will refer to as λ or the expected number of occurrences, is related to the binomial case by $\lambda = np$; however, there are many situations for which the sample size is never known, and one counts the number of failures, defects, etc., only. An example is the number of defects in a square yard of Quartermaster cloth and for which the standard or acceptable number may be only a single defect or even none.

For the sampling of a Poisson population, the sample size is set by specifying an acceptable expected number of occurrences λ_0 under the null hypothesis and an unacceptable number of occurrences λ_1 ($> \lambda_0$) under the existence of the alternative hypothesis. The approximate sample size n, determined very similarly to that for the binomial population by using asymptotically normal considerations, is

$$n = (1/4)\left(\frac{z_\alpha + z_\beta}{\sqrt{\lambda_1} - \sqrt{\lambda_0}}\right)^2 \tag{8-18}$$

and the critical region is based on

$$\tilde{\lambda} = \frac{z_\alpha\sqrt{\lambda_1} + z_\beta\sqrt{\lambda_0}}{z_\alpha + z_\beta} \tag{8-19}$$

In the determination of sample size the reader will no doubt understand that we have proceeded to control the errors of misclassification at two different values of the binomial or Poisson parameter, i.e., the acceptable one and the unacceptable one. We have not however made a computation of the entire OC curve or the power curve, but this calculation is rather easily performed. Note in particular that each of the Eqs. 8-5, 8-10, 8-16, and 8-18 may be solved for z_β in terms of the sample size n, the Type I error deviate z_α, etc. Thus for any values of these latter quantities, one may easily find the quantity z_β, which, when referred to a table of the standard normal distribution, will give the desired Type II error for that particular calculated condition. By changing these conditions one sees that the entire OC curve may be found and plotted if desired.*

As a brief summary of determining sample size for binomial and Poisson populations, we observe that the significance level type of approach is useful for the case in which one desires to estimate the population parameter to within a certain bound, whereas it seems of considerable value in practice to control errors of

*Many complete OC curves are given in Refs. 1 and 3.

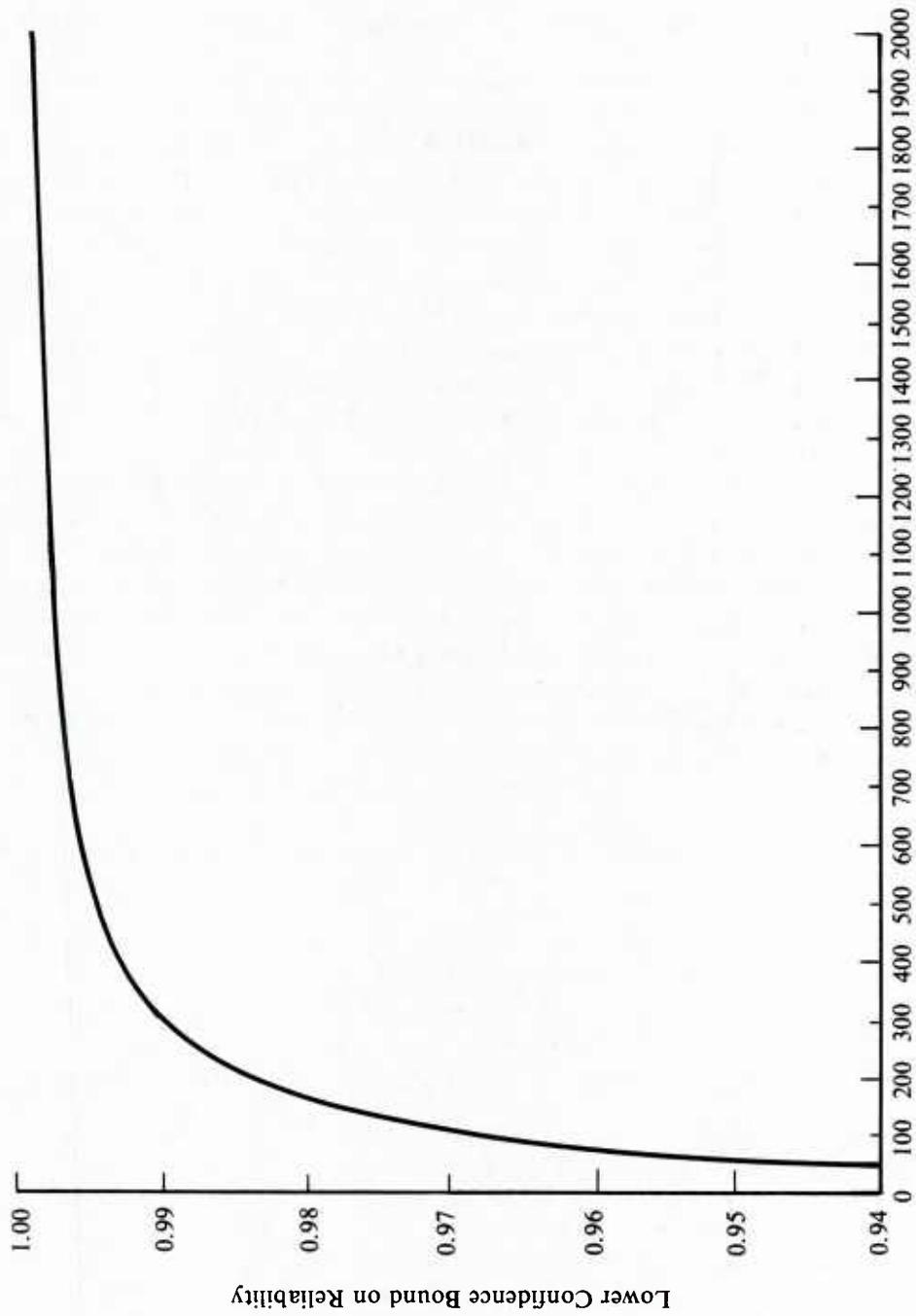


Figure 8-1. Number of Tests vs Improvement in Reliability

judgment in many applications. Nevertheless, for discrete data or binomial-type sampling, the sample sizes determined can be very large, unfortunately, for the desired protection. Perhaps a way around some of the difficulty in this connection is to perform some type of sequential sampling or, even better still, to try to inspect on a variables basis, which we discuss in the sequel. Finally, extensive binomial sampling to guarantee protection against critical defects would seem to lead invariably to very detailed evaluation or reevaluation of the basic design of the system.

Par. 8-3.2 discusses the comparison of two binomial populations.

8-3.2 SAMPLE SIZES TO COMPARE TWO BINOMIAL OR TWO POISSON POPULATIONS

When samples are drawn randomly from each of two binomial populations, we may no longer have primary interest in parameter estimation, but rather our interest centers around comparing the size of the two binomial parameters, which we will refer to as p_1 and p_2. In fact, very often we will have some rather key interest in knowing whether one of the p's is greater than the other or, just as importantly, whether the two different processes or treatments are equivalent, i.e., have equal p's. If we were primarily interested in the estimation problem, we could be better off to use the data of each sample to estimate the individual p's for the population from which each sample came. Alternatively, we would want to be quite sure that the two p's are equal, i.e., not significant in a statistical test, before we combined the two sample results for the purpose of estimating a common binomial parameter. Finally, we might have the problem or the desire to determine the required sample size that will lead to some control of results for estimating the common binomial parameter. Otherwise, we would be interested in detecting a given or stated difference between the two p's if one exists and is of significant practical interest or value.

Since we are now dealing with two binomial samples and we do not know whether they were drawn from a single binomial population, we must exercise some care in whether or not we estimate the variance on the basis of combining the two sample results into a single sample to estimate a common p or of keeping the two samples apart and thereby proceed as though we have distinct p's. This particular problem, as we recall from Chapter 5, was indeed the major consideration for comparing the two different binomial populations. Moreover, as we recorded in Chapter 5, a completely satisfactory answer to this point is still not available although it did seem best on practical grounds usually not to combine the two sample results but rather to treat the two p's as possibly being distinct. This consideration complicates the sample size determination problem somewhat although it is fortunately true that the arc sine transformation avoids the difficulty while it possesses rather good accuracy over the primary regions of interest. In fact, for p's very near zero or one the Poisson distribution can be used with excellent results.

We will frame the problem of determining sample size for comparing two binomial populations in terms of the following definitions:

p_1 = true unknown proportion of population designated as 1
p_2 = true unknown proportion of population designated as 2
$\hat{p}_i = x_i/n$ = occurrence ratio for estimating the parameter of the ith binomial population, $i = 1, 2$, with x_i the number of occurrences of interest
n = common sample size to be determined for the sampling of each binomial population.

For testing the hypothesis $H_0 : p_1 = p_2$ versus the possibility that $p_2 > p_1$ for the alternative H_1, the arc sine transformation leads to the approximate sample size of

$$n = (1/2)\left(\frac{z_\alpha + z_\beta}{\mathrm{Sin}^{-1}\sqrt{p_2} - \mathrm{Sin}^{-1}\sqrt{p_1}}\right)^2 \quad * \qquad (8\text{-}20)$$

If one were interested in guarding against whether p_2 is greater than or less than p_1, he would conduct a two-sided test with the significance level of $\alpha/2$ so that the desired overall level of the test would be α.

The critical region of the test (Eq. 8-20) is based on the quantity

*z_α and z_β are the upper α and β probability levels of $N(0,1)$.

$$z > \tilde{z} = z_\alpha / \sqrt{2n}. \tag{8-21}$$

Perhaps it would be illuminating to illustrate the two-binomial population sample size problem by referring to Example 5-3, in which a significant result was found. In Example 5-3 a combat simulation of Red versus Blue resulted in only six Blue infantrymen in 60 being lost in a battle versus 18 of 60 Reds; apparently, Blue seemed to have the superior rifle. In view of this, we will construct an example.

Example 8-5:

In a limited combat simulation between Blue and Red riflemen, it appeared that Blue might be able to kill about 30% of the Red riflemen, whereas Red's rifle capability was such that Red would kill only 10% of the Blue infantrymen. What sample size would be required to control errors of judgment to, say, 5%?

To solve the sample size problem here, we will set $p_1 = 0.10, p_2 = 0.30$, and use the one-sided test to be sure to pick up $p_2 > p_1$ if true and also take $\alpha = \beta = 0.05$. The calculation based on Eq. 8-20 gives $n = 81.4$ to control risks to 5% each; however, a sample size of 60 was used. (If we had set more liberal risks at, say, 10%, then an $n = 49$ would have been required. We see, therefore, that the determination of sample size in advance has some merit.)

Since we have given only the arc sine normal approximation for determining the sample size, we suggest that the reader may well use the significance test of Eq. 5-20, for which the variances are kept separate, and develop an asymptotic normal approximation for n. Once this is done, he should make a comparison of his calculation of n using the developed equation with the one we obtained by using the arc sine approach.

For the case of sampling two Poisson populations with parameters we will call λ_1 and λ_2, the sample size equation result is similar to the one for sampling a single Poisson population. In fact, the difference between the square roots of the mean number of occurrences is approximately normally distributed with the expected value equal to the difference in the square roots of λ_2 and λ_1, and the variance does not depend on the parameters but is equal to simply $1/(2n)$. Hence the sample size to control errors to the risks of α and β may be obtained from

$$n = (1/2)\left(\frac{z_\alpha + z_\beta}{\sqrt{\lambda_2} - \sqrt{\lambda_1}}\right)^2 \tag{8-22}$$

and the critical region depends on

$$z > \tilde{z} = 1/\sqrt{2n}. \tag{8-23}$$

Again, if one desires to plot the OC curve, he may solve Eq. 8-22 for z_β as a function of the other variables. Since many of the OC curves of this paragraph are based on asymptotic normality, their shape and general appearance would be similar to those of Fig. 6 of Ferris, Grubbs, and Weaver (Ref. 3) for the "normal test" and an equivalent sample size. Only OC curves needed for specific usage will be repeated in this chapter, however.

For binomial- and Poisson-type populations, therefore, we have given a number of useful equations to determine sample size and also have given a variety of redundant approximations to assure some accuracy of estimation. We believe that the procedures outlined herein should be sufficient for most applications the Army analyst will require in connection with sample size determination problems. For other cases the reader may extend his knowledge considerably by studying the references.

We have not covered the matter of determining sample sizes for general contingency tables in this paragraph since the problem here relates more to the use of the chi-square variate—a continuous random variable—and the ANOVA techniques. We will therefore proceed in par. 8-4 to discuss sample size determinations for continuous variates and initially will consider a treatment of the chi-square distribution and some of its applications.

8-4 SAMPLE SIZES FOR VARIANCE ESTIMATION AND COMPARISONS

8-4.1 SAMPLING A SINGLE NORMAL POPULATION TO ESTIMATE SIGMA

In Chapter 4 we discussed the sample variance, the sample standard deviation, and other measures of dispersion, such as the sample range, along with unbiased estimates of the normal population parameters. Also we established confidence bounds for appropriate population parameters. Since the chi-square distribution is more or less central to the statistical treatment of the sample variance and standard deviation, we describe some of its properties as related to the determination of sample sizes and the OC or power curves. In fact, it seems appropriate to deal first with either the variance or the standard deviation before proceeding with any treatment of mean values.

Although chi-square possesses a variety of applications to many different statistical problems, our initial discussion will involve the sampling of a normal distribution—either to obtain a "proper" estimate of the population variance or sigma or to control errors in assessing its size. In this connection, we recall that the quantity

$$\chi^2 = (n-1)s^2/\sigma_1^2 \tag{8-24}$$

with

$$s^2 = \Sigma(x_i - \bar{x})^2/(n-1) \tag{8-25}$$

follows the chi-square distribution with $(n-1)$ df. One should note in particular that σ_1 must be the standard deviation of the normal population actually sampled.

Our problem, loosely stated, is to determine the sample size necessary to estimate the true unknown normal population sigma. To do this, we may, as before, simply choose a significance level, such as the upper α probability level of the chi-square distribution, for which we would reject (with risk α) the null hypothesis if it is true and determine the sample size to obtain significance in case our null hypothesis may be false. Thus, and again, there is no effort to control the Type II error for a specified but very undesirable value of the normal population sigma. To be more specific and precise, especially in dealing with chi-square, one states that the σ_1 of the normal population is equal to a value σ, say, and thus his null hypothesis is $H_0: \sigma_1 = \sigma$. Then he calculates s^2 and substitutes these two values into Eq. 8-24 to obtain what we call the observed value of chi-square. This observed value is then compared with the selected significance level or percentage point of chi-square. We could be interested in whether the true unknown population sigma is much larger than, much smaller than, or just "different" than the hypothesized value we assign. Thus we would use, respectively, either an upper significance level only of chi-square, or only a lower percentage point, or judge whether the observed chi-square falls between the upper and lower levels selected to give a Type I error of α total for the two-sided type of significance test. We can see in this connection that it is wise to enumerate with specific symbols the exact percentage points to which we have referred. Since for the normal distribution the lower percentage points are the negative of the upper ones due to symmetry about zero, we have rather loosely called z_α the "upper" significance level when in fact a much improved and completely satisfactory designation would have been $z_{1-\alpha}$. Hence in dealing with the use of χ^2 we will call χ^2_α the lower percentage point and $\chi^2_{1-\alpha}$ the upper significance level. This means that for the two-sided test one would enter tables of the percentage points of chi-square with $\alpha/2$ and not α in order to have an overall level of α.

To proceed, we will now state that we want a high probability that if the true sigma of the normal population we actually sample is σ, we will accept this stated or null hypothesis. In fact, the rejection rate will be only α. Hence if we further specify that we want the observed s to have this chance of being no farther than some given distance from the hypothesized value $\sigma_1 = \sigma$ if true and we also want to guard against a normal population with a standard deviation much larger than our stated value of sigma, the form of our probability statement in percentage (fractional) change is

$$Pr[(s-\sigma)/\sigma \leq d] = 1 - \alpha \tag{8-26}$$

where

d = allowed fractional deviation from sigma.

By using R. A. Fisher's transformation of chi-square to approximate normality, which indicates that $(2\chi^2)^{1/2}$ is nearly normally distributed with mean $[2(n-1)]^{1/2}$ and variance unity, Thompson and Endriss (Ref. 14) have shown that the approximate sample size required is

$$n = z^2_{1-\alpha}/(2d^2).* \tag{8-27}$$

For the two-sided test for which one is interested in guarding against the sample standard deviation being too far below or too far above the true sigma of the normal population sampled, the upper half-alpha percentage point, i.e., $z_{1-\alpha/2}$, is used in Eq. 8-27. One can see that Fisher's transformation of chi-square is very useful indeed in this connection because it makes unnecessary a good bit of juggling around with the tables of percentage points of chi-square to determine the number of df by employing Eq. 8-24. (The sample size would then be one plus the number of df.)

We note that Eq. 8-27 is a very simple equation for determining the sample size because it requires only an upper percentage point of the standard normal distribution and the fractional (or percentage) deviation in terms of the unknown sigma allowed. (It does not consider Type II errors, however.)

Example 8-6:

A new conical boat-tailed artillery projectile—designed and developed for an 8000 m range—was thought to give a sigma in range of approximately 30 m, whereas current projectiles for this same firing condition were known to have a sigma in range of 45 m. Find the sample size needed for a verification test firing that would not allow the observed sigma to deviate more than, say, 15% above the desired value of 30 m with 95% assurance.

It is clear for this example that $\sigma = 30$, $d = 15\%$, and $\alpha = 0.05$. Hence we see from Eq. 8-27 that the sample size n is determined from

$$n = (1.645)^2/[2(0.15)^2] = 60.13 \text{ or } n = 60.$$

In summary, therefore, if we fire 60 rounds of the newly proposed projectile and compute its standard deviation in range, we would have 95% assurance that if the true sigma were indeed 30 m, the observed sigma would exceed this value by more than (0.15) (30) = 4.5 m. (If the true sigma were much greater than 30, Eq. 8-24 likely would show significance.)

We remark for this example that we have depended only on the idea of establishing significance if it be the case, so that the sample size is determined without consideration of placing a low risk on the possibility that the new projectile may even have a sigma equal to that of the current projectile. We will, therefore, now consider this other method of determining n and make a comparison of the two. Is there better guidance than planning to use a sample as large as 60?

For the control of errors of the misclassification approach, we set the null and alternative hypotheses as follows:

$$\begin{aligned} \text{Null hypothesis:} \quad & H_0\text{: } \sigma_1 = \sigma, \text{ with rejection risk of } \alpha \\ \text{Alternative hypotheses:} \quad & H_1\text{: } \sigma_1 = \lambda\sigma, \text{ with } \lambda > 1 \text{ and varying } \beta. \end{aligned}$$

Thus for this particular formulation we are using a one-sided test and in particular are guarding against a larger sigma than we can tolerate in our decision. For this case Ferris, Grubbs, and Weaver (Ref. 3) have shown that the ratio of the undesirable sigma to the stated value of sigma and the probability levels of chi-square are functionally related as follows:

$$\lambda = (\chi^2_{1-\alpha}/\chi^2_{\beta})^{1/2} \tag{8-28}$$

*$z_{1-\alpha}$ is the upper α probability level of the standardized normal distribution.

where the number of df for chi-square is understood to be $(n-1)$, and we fix the Type I error rate but allow the Type II error rate β to vary. Thus for all sample sizes and any value of λ, the Type II errors can be found or β may be taken as some percentage point and the value of λ determined so that the entire OC or power curve may be obtained. The OC curves of the chi-square test based on Eq. 8-28 are given as Fig. 8-2, which is a repeat of Fig. 4-1 of Ref. 1. We illustrate the use of Fig. 8-2 in Example 8-6.

For the purpose of detecting a normal population sigma much less than that hypothesized, the OC curves are given here as Fig. 8-3, which is Fig. 4-2 of Ref. 1.

Again, some juggling is required to obtain very clear-cut answers from Eq. 8-28 since it does not give the sample sizes directly. Some approximate equations for determining the sample size directly can be given, however, and the first one we list is based on the assumption that the sample standard deviation is nearly normally distributed. This is not a really "wild" assumption; it has long been more or less "accepted" that the chi-square distribution is "nearly normal" when the df are "approximately thirty or more". Moreover, to have any respectable power in making any important decisions, one can probably expect the sample sizes must be about 25 or more! With this assumption and by applying the rather general expression (Eq. 8-5) to this case, it can be shown that the approximate sample size to control Type I and Type II errors to α and β is

$$n = (1/2)\left(\frac{z_\alpha + z_\beta \lambda}{\lambda - 1}\right)^2 * \tag{8-29}$$

The critical region (Ref. 4) is

$$z > \tilde{z} = \frac{\lambda\sigma(z_\alpha + z_\beta)}{z_\alpha + \lambda z_\beta} \tag{8-30}$$

Another approximate equation for the sample size is given by Chand (Ref. 4) and is based on using the distribution of $\ln(s^2)$, which has been shown by Bartlett and Kendall (Ref. 15) to be more nearly normally distributed than s^2, and furthermore, this logarithmic variance has the desirable property that its distribution depends on the unknown population sigma only in its expected value. The approximate sample size based on the logarithmic variance is

$$n = 1 + 2\left(\frac{z_\alpha + z_\beta}{\ln\lambda^2}\right)^2 \tag{8-31}$$

Based on a comparison of Eq. 8-29, Eq. 8-31, and the more exact values that may be determined with the aid of Eq. 8-28, Chand (Ref. 4) has shown that all three estimates of the sample size are only a very few, if any, units apart, and for the cases considered the agreement is within a unit. Thus it seems safe to conclude that a very satisfactory determination of the sample size to control errors of misclassification in tests of hypotheses about the size of the normal population variance or sigma can be obtained by any of the three methods. Let us now give an example (Example 8-7) that brings out some of these points.

Example 8-7:

Referring to Example 8-6, let us now add the condition that we would like to be able to reject the null hypothesis that sigma is equal to 30 m with 95% assurance if in fact it were the same as that of the present round, i.e., 45 m.

It is now clear that we have $\lambda = 45/30 = 1.5$ and $\beta = 0.05$ in addition to the basic data of Example 8-6. Referring to Fig. 8-2 for $\lambda = 1.5$, we see that the required sample size for the desired protection is no more than approximately $n = 35$, if that large, as we read the curves. On the other hand, if we were to calculate n from Eq. 8-29, we would get $n = 33.8$, and the calculation based on Eq. 8-31 gives $n = 33.9$. These sample sizes are

*One must be careful to note that Chand's λ in Ref. 4 is actually the square of ours.

certainly close together on practical grounds and are smaller than the $n = 60$ calculated without regard to the control on the Type II error. In fact, a sample of size $n = 60$ almost surely would reject the null hypothesis that the round-to-round sigma in range is 30 m when it is actually 45 m. Moreover, by reconstructing the problem somewhat differently, one may show by using Eq. 8-29 or Eq. 8-30, that for a sample size of 60 the Type I and Type II errors (or "producer" and "consumer" risks) may be reduced to practically negligible values $\alpha = \beta = 0.007$. Thus we see that the sample size of $n = 60$ may not be needed for this particular problem and that on practical grounds it seems best to set acceptable and rejectable values of the unknown population sigma with suitable risks to determine sample size. Moreover, as the sample size increases, there seems to be little justification for sticking with a Type I error as high as 0.05 when this risk could probably be reduced to a lower value such as 0.01, etc. Thus it appears to be wise to frame the sample size problem very carefully in terms of the practical problem.

With regard to Example 8-7, it will be of some interest for the reader to use Fig. 8-3 to find the sample size that will detect a sigma of 30 m when it is hypothesized that the sigma of the normal population sampled is 45 m, the larger value.

8-4.2 CHI-SQUARE SAMPLE SIZES FOR CONTINGENCY TABLES OR FOR CURVE FITTING

Since the chi-square distribution is very widely used or is found to solve many diverse problems in statistics, it should be expected that the chi-square statistic may be employed to estimate sample sizes for contingency tables or for the fitting of frequency curves to show a good or poor fit, etc,. Thus expressions such as Eq. 8-28 are found to be much more general in application than thought initially because the association relates the power of a significance test for the parameters involved. In fact, as an example and alternate derivation, the reader may substitute Fisher's transformation of chi-square in Eq. 8-28 and show that this will lead directly to Eq. 8-29 for sample size.

In the statistical analysis of contingency tables as presented in Chapter 5, one often will want to know whether his sample size is "adequate" or, better still, will try to plan his experiment in advance by using the proper sample size at the beginning. If one has some preliminary data on observed proportions for a contingency table study that will be carried out and knows fairly well the expected or theoretical proportions, the sample size may be estimated from

$$n = \chi^2 / \left(\sum_{i=1}^{k} p_i^2 / P_i - 1 \right) \tag{8-32}$$

where

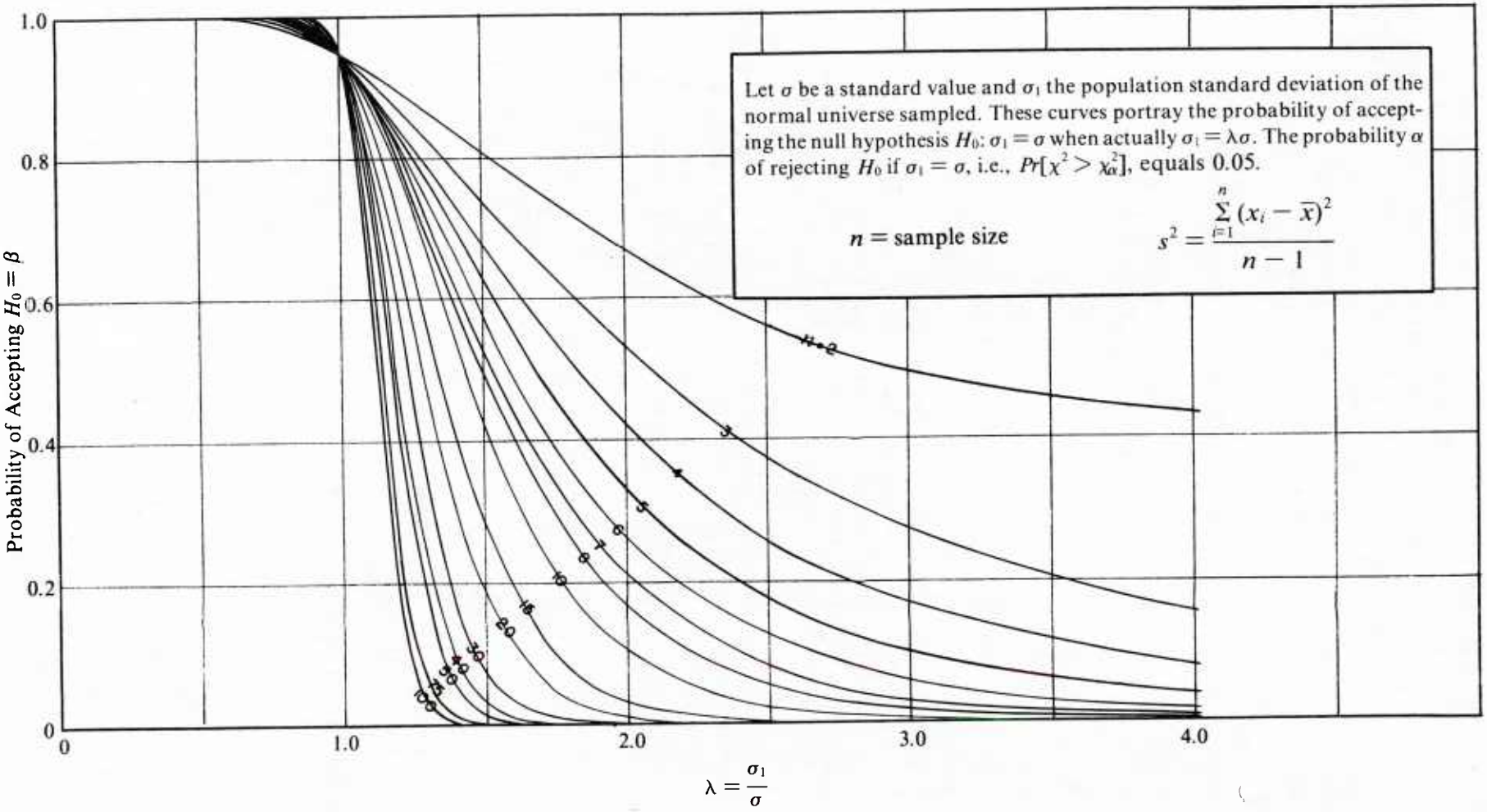
p_i = preliminary observed proportion
P_i = expected or theoretical proportion
k = number of classes in the contingency table
χ^2 = upper or lower significance level of chi-square.

If one is dealing with frequencies instead of proportions, the sample size may be determined from

$$n = \sum_{i=1}^{k} f_i^2 / F_i - \chi^2 \tag{8-33}$$

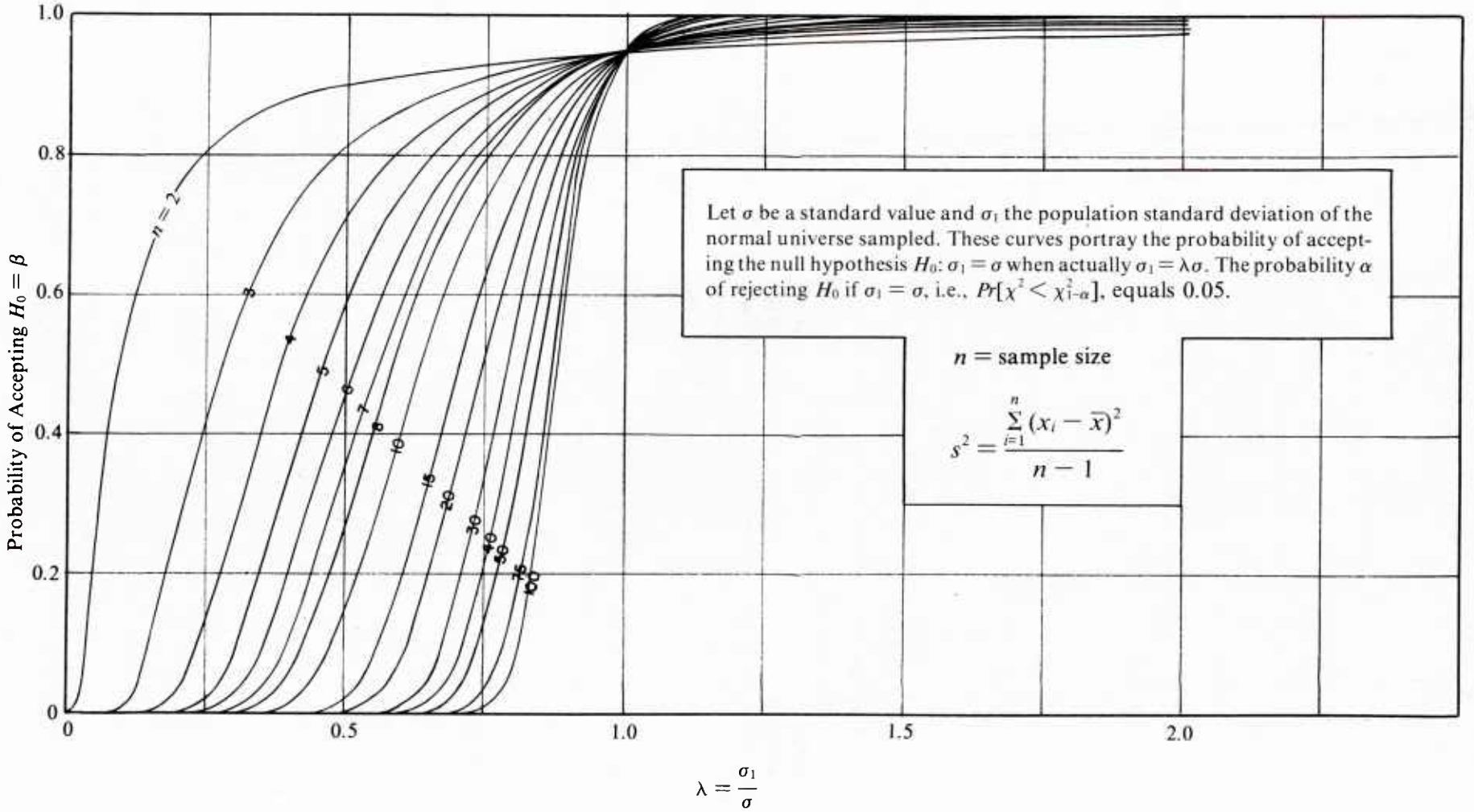
where

f_i = preliminary or observed frequency for the ith class
F_i = theoretical frequency
χ^2 = some percentage point, e.g., the 95% point.



Reprinted with permission from "Operating Characteristics for the Common Statistical Tests of Significance" by Charles D. Ferris, Frank E. Grubbs, and Chalmers L. Weaver, Annals of Mathematical Statistics XVII, No. 2 (June 1946). Copyright © by Institute of Mathematical Statistics.

Figure 8-2. Operating Characteristics of the χ^2-Test $\left[\chi^2 = \frac{(n-1)s^2}{\sigma^2}\right]$ for Testing $\sigma_1 = \sigma$ Against $\sigma_1 > \sigma$ (Ref. 3)



Reprinted with permission from "Operating Characteristics for the Common Statistical Tests of Significance" by Charles D. Ferris, Frank E. Grubbs, and Chalmers L. Weaver, Annals of Mathematical Statistics XVII, No. 2 (June 1946). Copyright © by Institute of Mathematical Statistics.

Figure 8-3. Operating Characteristics of the χ^2-Test $\left[\chi^2 = \frac{(n-1)s^2}{\sigma^2}\right]$ for Testing $\sigma_1 = \sigma$ Against $\sigma_1 < \sigma$ (Ref. 3)

Uitert (Ref. 16) suggests that an alternate form for estimating the sample size n is

$$n = n_a \chi_d^2 / \chi_a^2 \tag{8-34}$$

where

χ_a^2 = "available" value of chi-square from preliminary data
n_a = number of observations on which χ_a^2 is based
χ_d^2 = "desired" or projected significant value of chi-square.

In passing, we remark that Eq. 8-32 and/or Eq. 8-33 when solved for chi-square will give very useful methods of computing χ^2. See, for example, Allison (Ref. 17).

Example 8-8:

In Example 5-5, which represented a "double dichotomy" type of contingency table analysis, it was found that of 40 recruits selected at random and divided into one group of 18 who had previous experience shooting rifles and a second group of 22 who did not have any rifle-shooting experience, no discernable difference in expertise was shown in rifle practice. In fact, the observed proportions of 12 in 18 showing expert and 9 in 22 showing the same degree of proficiency could occur by chance about 10% of the time under the null hypothesis of no difference. Could it be that the sample size was too small, and if so, what sample size would be suggested for another test since there seems to be a "practical" difference in the two ratios?

Although chi-square was not calculated in Example 5-5, from Eq. 5-11 it is

$$\chi^2 = 40[(12)(13) - (6)(9)]^2/[(18)(22)(21)(19)] = 2.63$$

with 1 df. We note from a table of percentage points of chi-square with 1 df that an observed value of chi-square equal to about 3.85 would have been significant at the 95% level. Hence we note from Eq. 8-34 that

$$n = 40\,(3.85)\,/\,(2.63) = 59$$

or that is, if we were to run another experiment, it would be wise on the basis of this evidence to test 59 or more recruits, about half with and half without experience shooting rifles.

8-5 SAMPLE SIZES FOR COMPARING TWO NORMAL POPULATIONS VARIANCES

The variance-ratio test or the Snedecor-Fisher F statistic, which is the ratio of two sample variances, is used to test the hypothesis that the true variances of two normal populations are equal. This significance test is often carried out as a preliminary test before Student's t statistic is applied to compare normal population means (Chapter 4). If the two normal populations sampled have unequal variances, this should be known to the experimenter. Thus one would show some concern if the variance of one population were much larger than that of the other, and he would like to settle this point early. Moreover, if we are going to conduct the variance-ratio test, it is appropriate to have the proper sample size. Therefore, to study the problem of sample size determination, we define the following:

σ_1 = true unknown standard deviation of the first normal population
σ_2 = true unknown standard deviation of the second normal population
n_1 = sample size for sampling the first normal population
n_2 = sample size for sampling the second normal population
s_1^2 = sample variance based on $(n_1 - 1)$ df for the first sample
s_2^2 = sample variance based on $(n_2 - 1)$ df for the second sample
$\lambda = \sigma_1/\sigma_2$ = ratio of the true unknown standard deviation of the first population to that of the second.

In a manner similar to that of finding the sample sizes for the previous significance tests, we could determine the sample size from being able just to detect significance should it occur. We might, on the other hand, proceed to find the sample size to control the error of rejecting the null hypothesis if the variances are equal but to be relatively sure of rejecting this hypothesis if the quantity λ should be as large as, say, 1.5 or 2, for example. We might say, however, that the present problem is a bit different from the preceding ones of this chapter. Specifically, we are not trying "to get close to" a parameter of the single population we are sampling; rather our prime interest centers around learning as much as possible about the ratio λ of the two unknown population standard deviations from available data. Of particular interest, for example, is the determination of the sample size such that the ratio of the two population sigmas will be within confidence limits of a given range. (The problem of placing confidence bounds about the ratio of the two sigmas was covered in Chapter 4 but not necessarily from the standpoint of sample size determination.)

If the two sigmas were actually equal, i.e., $\lambda = 1$, the F statistic defined by

$$F = s_1^2/s_2^2 \tag{8-35}$$

would follow the Snedecor F distribution exactly. On the other hand, for the case of unequal sigmas this is not so although the quantity given by

$$F = s_1^2\sigma_2^2/(s_2^2\sigma_1^2) = s_1^2/(\lambda^2 s_2^2) \tag{8-36}$$

which has been corrected for the ratio of sigmas, does follow F. Moreover, it should be clear that the relation (Eq. 8-36) enables one to determine the power function or the OC curve of the F test rather easily. In fact, Ref. 3 shows that the relationship between the percentage points of the F statistic with $(n_1 - 1)$ df in the numerator and $(n_2 - 1)$ df in the denominator, and the ratio λ of the two unknown sigmas is

$$\lambda = (F_{1-\alpha}/F_\beta)^{1/2} \tag{8-37}$$

where

$F_{1-\alpha}$ = upper α probability level of F
F_β = lower β probability level of F.

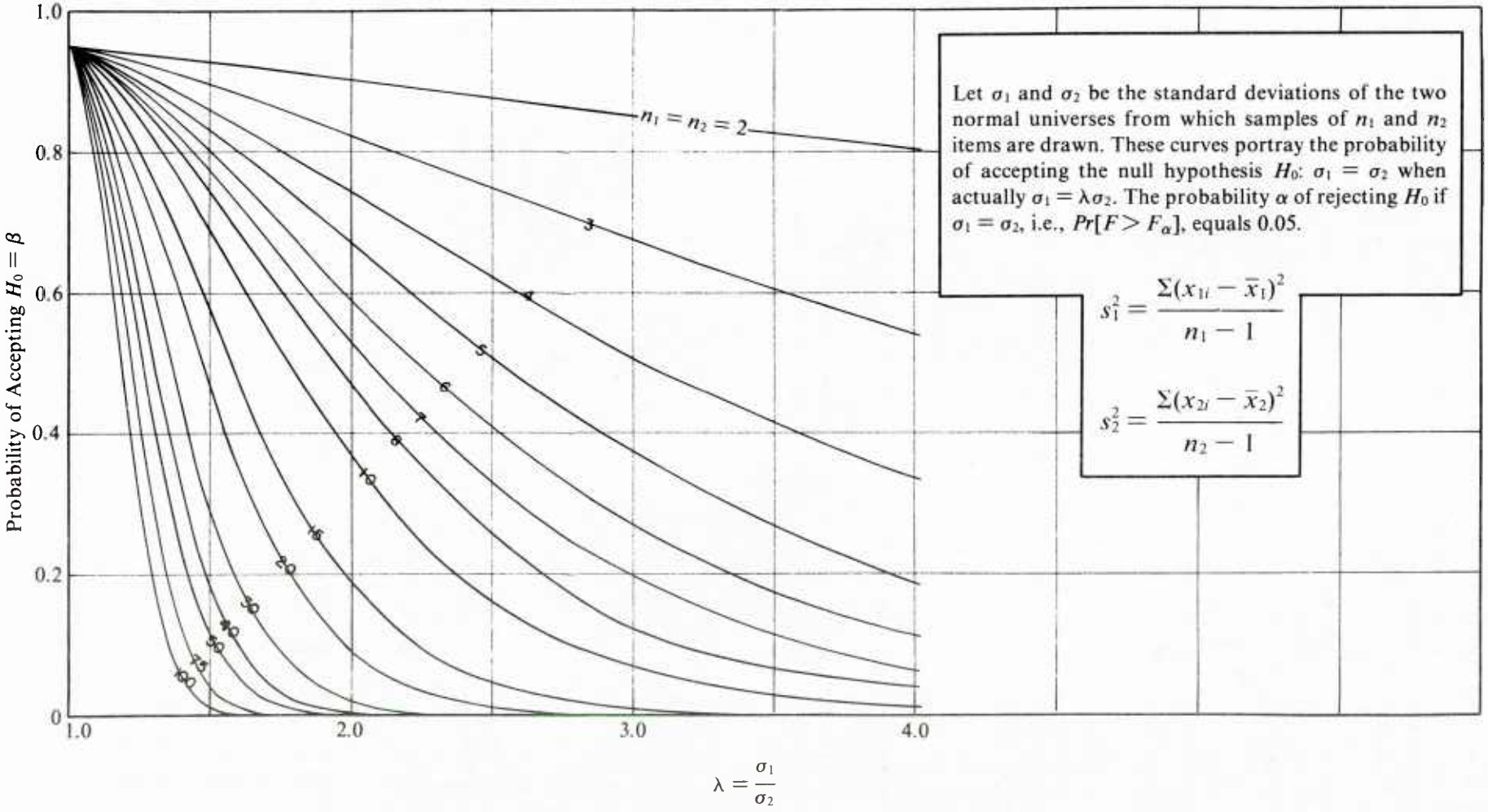
Hence with the aid of Eq. 8-37 and tables of the percentage points of F, one can plot the OC curves of the F test, which we give in Fig. 8-4 for equal sample sizes. (Fig. 8-4 is taken from Ref. 3 and may be found also in Ref. 1. For unequal sample sizes, or unequal df in the numerator and denominator of F, OC curves are given in Refs. 1 and 3, which originally were published in Ref. 3.) It should be noted that Fig. 8-4 is only for a Type I error of 0.05. To find the sample size from Fig. 8-4, one also must specify the Type II error he is willing to accept and the objectionable ratio of sigmas, so that with λ he enters the curves and reads the sample size n for the value β on the ordinate scale. We illustrate this in Example 8-9, but first we take up the matter of suitable equations to calculate the sample sizes directly—at least for many applied problems.

Eq. 8-37 does not lend to the calculation of the sample size n in a very direct manner. Nevertheless, by using Fisher's Z, which is related to the Snedecor F through

$$Z = (1/2)\ln F \tag{8-38}$$

and which is nearly normally distributed for large enough df, it can be shown without much difficulty that the approximate relationship between the sample sizes, the ratio λ, and the standard normal percentage points is

$$2\left(\frac{1}{n_1 - 1} + \frac{1}{n_2 - 1}\right) = [\ln\lambda^2/(z_{1-\alpha} + z_\beta)]^2 . \tag{8-39}$$



Reprinted with permission from "Operating Characteristics for the Common Statistical Tests of Significance" by Charles D. Ferris, Frank E. Grubbs, and Chalmers L. Weaver, Annals of Mathematical Statistics XVII, No. 2 (June 1946). Copyright© by Institute of Mathematical Statistics.

Figure 8-4. Operating Characteristics of the F-Test $\left[F = \frac{s_1^2}{s_2^2}\right]$ for Testing $\sigma_1 = \sigma_2$ Against $\sigma_1 > \sigma_2$ $(n_1 = n_2)$ (Ref. 3)

Chand (Ref. 4) points out that Eq. 8-39 is not very accurate for sample sizes as low as about six although, in practical applications, one might expect to have to deal with much larger sample sizes. In any event, with the aid of Eq. 8-39 one can substitute values of n_1 and n_2 on the LHS of Eq. 8-39 until a match in the calculated value is attained with the right-hand side (RHS). Furthermore, when the sample sizes are equal, i.e.,

$$n_1 = n_2 = n \tag{8-40}$$

the common sample size for the variance-ratio F is found to be approximately equal to

$$n = 1 + 4\,[(z_{1-\alpha} + z_\beta)/\ln\lambda^2]^2 \tag{8-41}$$

so that the sample size per variance is about double that for the chi-square test as in Eq. 8-31 in sampling a single normal population to estimate the population sigma, and for the same risks.

A slightly different approach to estimate sample sizes for pinning down the ratio of two normal population sigmas is to use an approximation for F that depends on a sufficiently "large" number of df and, hence, may be no real problem. This rule states that when one of the df is fairly large, the F ratio can be constructed so that F is nearly distributed as chi-square divided by the numerator number of df. (We should state here that the "textbook" rule to place the largest sample variance in the numerator of F is rather artificial—and perhaps even a bit confusing or misleading—for actually one may take the ratio in the practical order of variances desired, especially since the lower percentage points of F may be found by switching the numbers of df and taking the reciprocal of the F so obtained to find the correct percentage points anyway!) This particular transformation of F to an approximate chi-square would lead to the sample size of

$$n = \left(\frac{z_\alpha + \lambda z_\beta}{\lambda - 1}\right)^2 \tag{8-42}$$

which is clearly double that of Eq. 8-29 for the variance estimation problem in sampling a single normal population. In fact, the reader may examine Figs. 8-2 and 8-4 simultaneously in this connection. He will note that if he enters Fig. 8-2 with any value of λ and goes to the sample size curve for selected value of the probability of accepting H_0, he will find that the sample size so determined is only about one-half that for the same λ and acceptance probability on Fig. 8-4 for the F test of the ratio of population sigmas. Thus it may be remarked that for practice one could get by quite well with only the chi-square OC curve of Fig. 8-2!

Example 8-9:

Let us return to Example 4-5 concerning the firing of only ten 20-mm projectiles for which the variance in the horizontal direction was compared with that in the vertical direction by using the F test. It was found that no significant difference was observed in the horizontal and vertical true sigmas. What sample size would one need to fire to reject the hypothesis of equal sigmas with 95% probability if the true sigma in the vertical direction were actually 1.5 that of the horizontal true sigma? By entering Fig. 8-4 with $\lambda = 1.5$ and by trying to read the OC curves for a β probability of 0.05 on the ordinate scale, we see that the sample size n is greater than 50 but less than 75. A computation using either Eq. 8-41 or Eq. 8-42 gives an $n = 67$. Thus the test of only 10 rounds becomes somewhat superficial, and a much larger sample size would have been required to pick up even a 50% difference in the horizontal and vertical sigmas!

So far we have used the power function of the Snedecor-Fisher F test to determine sample sizes for the comparison of two normal population sigmas or to control the ratio of them. However, we should remark, as is well-known, that the F variance ratio is much more general in application. In fact, the F ratio is just as important in the ANOVA test for any number of treatments, and thus we would often need to determine sample sizes here. We will reserve this type of discussion for a later paragraph. It is best now to proceed with sample size determination problems for one or two populations.

With our discussion of the problem of sample size estimation to compare two unknown normal population sigmas, we are now ready to take up the next topic, i.e., sample size determination for normal population means.

8-6 SAMPLE SIZES FOR ESTIMATION OF NORMAL POPULATION MEANS

8-6.1 SAMPLE SIZES FOR MAKING INFERENCES ABOUT THE SIZE OF A NORMAL POPULATION MEAN

The idea is to draw a single random sample of size n from some normal population and on the basis of it to determine the size of the true mean within given bounds. The ordinary Student's t test is a natural statistic for this purpose since the only unknown population parameter in it is the population mean itself. Moreover, the t statistic has the sample size directly in it! As we well know, Student's t for a single sample from $N(\mu,\sigma)$ is

$$t = \frac{\sqrt{n}(\bar{x} - \mu)}{s} \tag{8-43}$$

and this quantity follows the t distribution with $(n - 1)$ df as in Eq. 4-100. Suppose we would like to require with "large" probability that the population mean will be within a given distance of the sample mean or to be able to pick up a departure, say, d between the two if it occurs. That is, we want

$$\begin{aligned} &Pr[-d \leq (\bar{x} - \mu) \leq +d] = 1 - \alpha \\ &= Pr[-\sqrt{n}d/s \leq \sqrt{n}(\bar{x} - \mu)/s \leq \sqrt{n}d/s]. \end{aligned} \tag{8-44}$$

But since the middle quantity is t and hence distributed as Student's t, we can equate the positive bound to the upper half-alpha level of probability and solve for the sample size n from

$$n - 1 = \frac{s^2 t^2_{1-\alpha/2}}{d^2} - 1.* \tag{8-45}$$

Thus the sample size necessary to guard against a departure of the population mean from the sample mean by as much as d or to detect the departure d if it should occur is determined from the sample variance multiplied by the square of the half-alpha probability level of Student's t divided by the square of the departure sought. In this connection, the reader should note that we have not assumed that the true sigma of the normal population sampled is known. Rather, we may have merely an estimate s of it. Had we actually known the true sigma, we could simply replace s in Eqs. 8-43 through 8-45 with it and deal with a normally distributed statistic instead of a t variate, and the sample size would then be determined in terms of the known population variance in Eq. 8-45.

Another possibility for this type of problem is to hypothesize that the true mean of the normal population sampled is, say, $\mu = a$ and to compute the t of Eq. 8-43 as if this were so. However, should it be that the correct value μ of the true mean of the normal population departs from a by the amount d, on the average $(\bar{x} - a)$ would either increase or decrease by the amount d; therefore, we would have confidence $(1 - \alpha)$ that such deviation would be noticed in our significance test.

In summary, our test procedure is simply to be able, with "high confidence", to observe some departure d in means if it occurs, and we have set only the Type I risk level but not the Type II error, which we might like to guard against also.

An alternate, approximate procedure for sample size determination is to divide the sum of squares (SS) about the sample mean by $(n - 3)$ instead of the usual $(n - 1)$ and hence have a quantity that is almost normally distributed, as in Eq. 4-105. This new quantity will be referred to as z, a normally distributed variate, so that the relation with t is given by

$$z = t[(n - 3) / (n - 1)]^{1/2}. \tag{8-46}$$

*This equation has df on the left since Student's t is in df.

This would mean that the sample size n is determined from

$$n = \frac{s_a^2 z_{1-\alpha/2}^2}{d^2} \tag{8-47}$$

where the new variance s_a^2 is based on the divisor $(n-3)$. It will be of interest to make a comparison of these two methods, or equations, for estimating sample size.

Example 8-10:

Given the 11 muzzle velocities of Example 4-1, use these data to find the sample size necessary to determine the true muzzle velocity of the 155-mm projectiles within a distance of one sample standard deviation with 95% confidence.

The sample standard deviation in Example 4-1 based on $(n-1)$ df is 10.25, so that we take $d = 10.25$, which cancels with the s of Eq. 8-45 anyway. Hence for the 95% confidence level all we have to do is look in a table of the 97.5% points of Student's t until the square of a value of t in this column minus one equals the number of df or the square root of one plus the number of df is equal to the tabulated point. We find for this problem that $(n-1)$ is just larger than 5, so that we would take $n = 6$, which is a smaller sample size than in Example 4-1.

Alternatively, s_a based on the divisor $(n-3)$ instead of $(n-1)$ would be 11.46 instead of 10.25; therefore, the sample size calculated from Eq. 8-47 would be about 5. Since we are dealing with very low sample sizes, it cannot be expected that the agreement, especially with an approximation, would be perfect.

Thus in drawing a random sample from a single normal population, we see from the examples that it will often be of interest to decide just what we are really sampling for, especially since we may be able to save on costs of tests that otherwise might be expensive. The proper determination of sample size often leads to some surprising conditions in experimentation! Again, however, we have so far dealt with only one end of the OC curve in our test relative to a normal population mean. We say this even though we do have a useful procedure for assuring a high degree of confidence that if a difference of interest is present we will notice it, except for a low chance result. Let us now, however, proceed to the use of the entire OC curve for Student's t statistic or at least to the two key points for the size of an acceptable population mean and the alternative relative to an objectionable or unacceptable value of the mean. As before, we frame the problem in terms of a null and an alternative hypothesis.

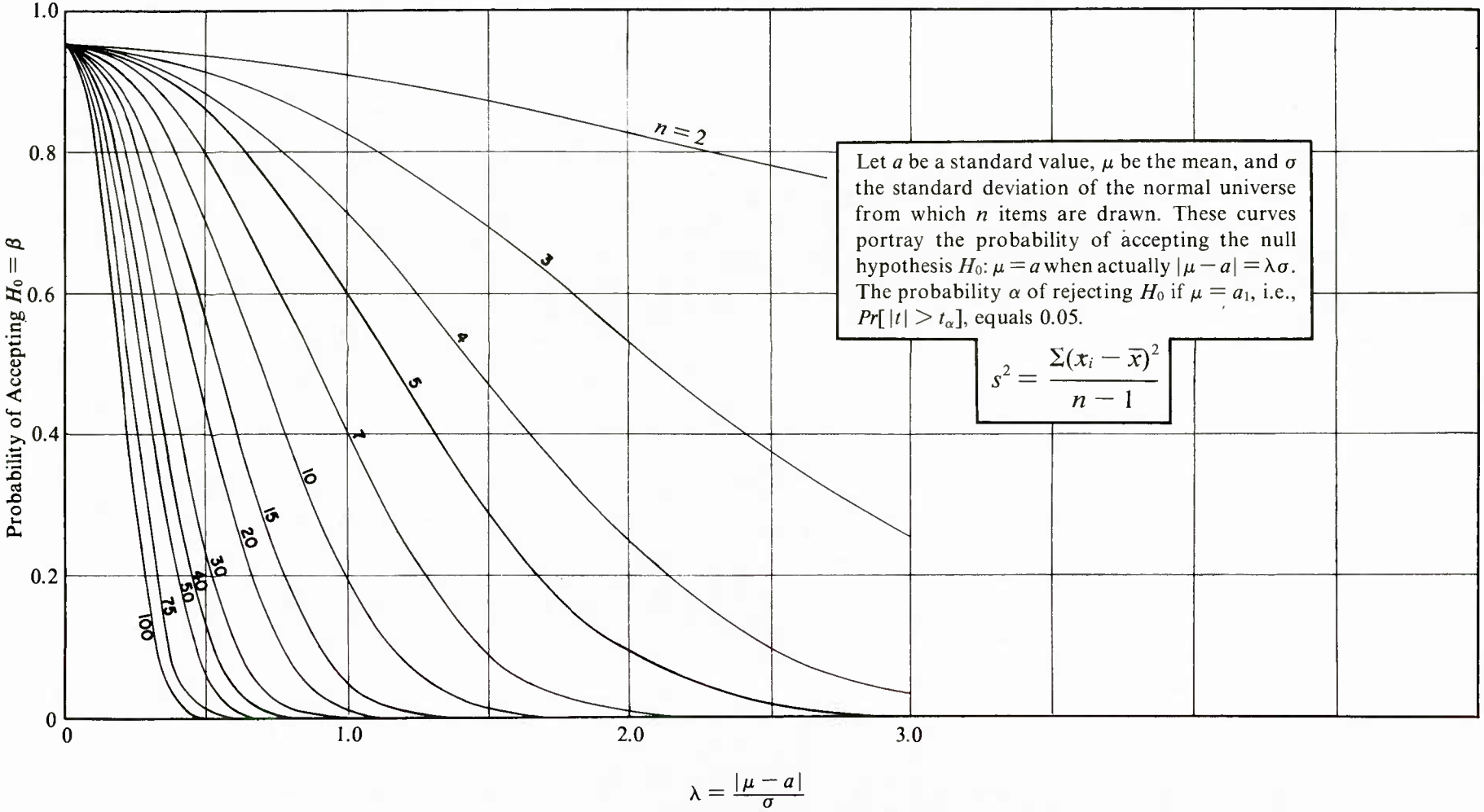
For the control of Type I and Type II errors approach, the null hypothesis is

$$H_0: \text{the unknown normal mean } \mu = a.$$

Then we will be concerned to determine the sample size to guard against alternatives of the form

$$H_1: |\mu - a| = \lambda\sigma$$

that is, if the departure of the true mean μ from our hypothesized value a is some lambda sigma units away, we will want to be able to detect this with high probability, especially if lambda is, say, as large as 1.5 or 2. The reader should note in particular that we have expressed the deviation in units of sigma since this seems to be desirable and indeed also fits in better with the theory. OC curves for the t test were published in 1946 by Ferris, Grubbs, and Weaver (Ref. 3) as their Fig. 7. These curves are also given in Ref. 1 and repeated here as Fig. 8-5. The abscissa of Fig. 8-5 is for values of the relative deviation λ in the number of sigma units the true mean is from the stated or hypothesized mean value, and the ordinate gives the chance of accepting the null hypothesis of no difference as a function of the quantity lambda. The Type I error is 0.05 only. As a quick example, suppose one desired the sample size to be able to detect with 95% assurance a departure of the true normal mean from the stated value of one sigma. Then, by entering the curves of Fig. 8-5 with $\lambda = 1$, he will read that $n \approx 15$. Thus one would select a sample of size 15 from the single normal population, carry out Student's t test at the upper 5% level, and reject the null hypothesis of no difference if the observed t exceeds the 95% level of t for 14 df.



Reprinted with permission from "Operating Characteristics for the Common Statistical Tests of Significance" by Charles D. Ferris, Frank E. Grubbs, and Chalmers L. Weaver, Annals of Mathematical Statistics XVII, No. 2 (June 1946). Copyright© by Institute of Mathematical Statistics.

Figure 8-5. Operating Characteristics of the t-Test $t = \left[\frac{\sqrt{n}(x - a)}{s}\right]$ for Testing $\mu = a$ Against $\mu \neq a$ (Ref. 3)

If the null hypothesis H_0 is true, the chance of accepting it is given by

$$Pr[-t_{\alpha/2} \leq t \leq +t_{\alpha/2}] = 1 - \alpha \tag{8-48}$$

where we usually set $\alpha = 0.05$. On the other hand, if H_0 is not true and some alternative H_1 becomes true because the correct mean of the normal population sampled is not equal to a, but departs from it by lambda units of the population standard deviation σ, Eq. 8-48 becomes equal to β, where from Ref. 3,

$$\beta = Pr[-t_{\alpha/2}s/\sigma + \lambda\sqrt{n} < \sqrt{n}(\bar{x} - \mu)/\sigma < +t_{\alpha/2}s/\sigma + \lambda\sqrt{n}] \tag{8-49}$$

and also where λ is equal to

$$\lambda = |\mu - a|/\sigma. \tag{8-50}$$

Thus for any given values of the percentage points of the t distribution selected—along with values of sigma, the deviation λ in sigma units, and the sample size—one can calculate the power or OC curves from Eq. 8-49. Ref. 3 covers several methods that were used and checked against each other to determine the OC curves given on Fig. 8-5 for the t test. In fact, instead of the form given in Eq. 8-49, Ref. 3 shows that one may use the chance of Type II errors β as expressed by

$$\beta = Pr\left[-t_{\alpha/2} < \frac{\sqrt{n}(\bar{x} - \mu)/\sigma + \lambda\sqrt{n}}{s/\sigma} < +t_{\alpha/2}\right] \tag{8-51}$$

where in the middle expression of the numerator, the first term is a unit normal variable and the second term is known as the noncentrality parameter of the noncentral t statistic expressed by the middle fraction. The denominator in the middle term is a chi variate. Rather extensive tables of the noncentral t distribution were published in 1957 by Resnikoff and Lieberman (Ref. 18). With their tables the OC or power curves may be determined for Student's t statistic.

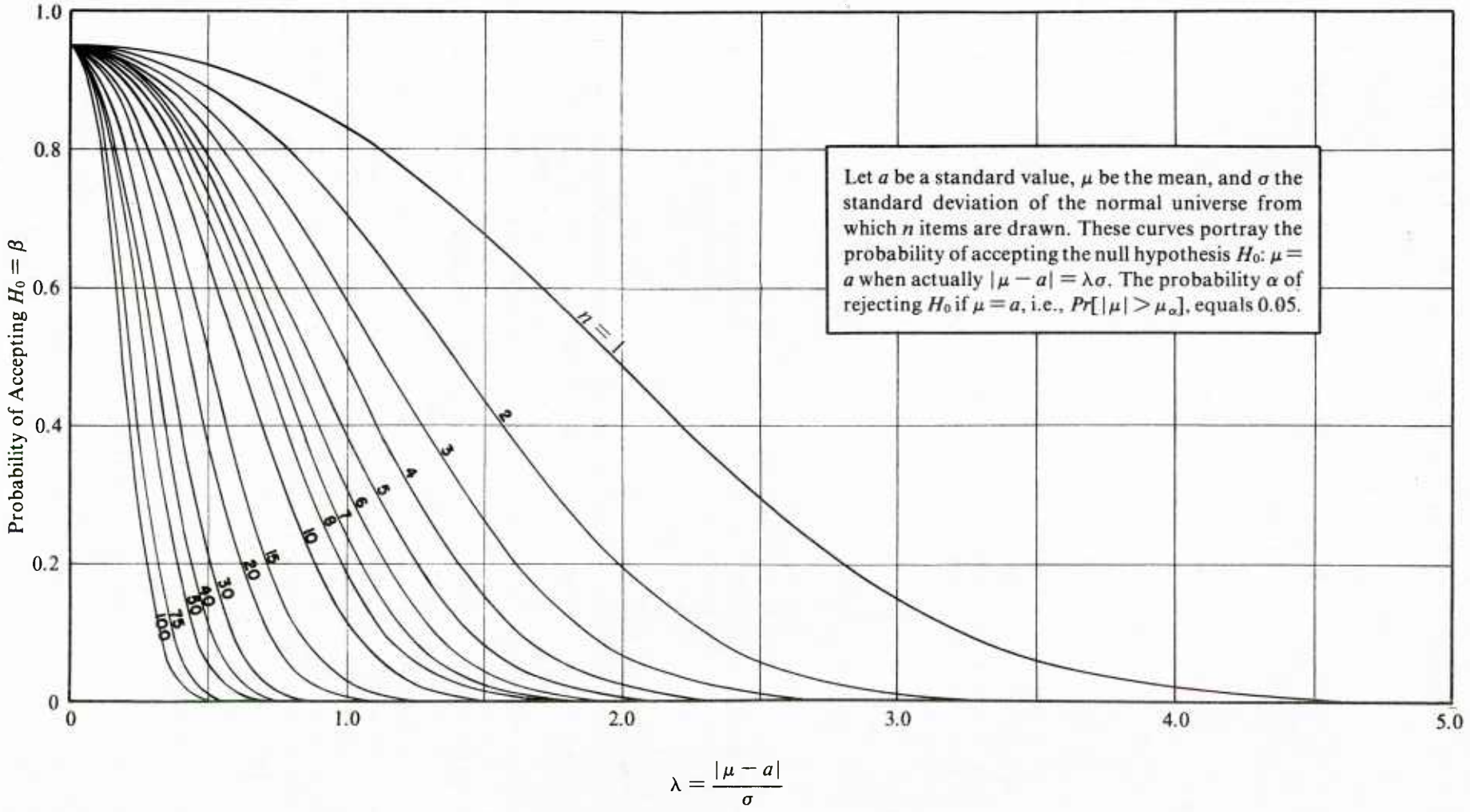
In comparison with the OC curves of Student's t on Fig. 8-5, we give on Fig. 8-6 the OC curves of the normal test, which assumes that sigma is known as indicated. Of course, the two sets of OC curves are very similar, and as the sample size increases we know that t becomes normally distributed so that ultimately the OC or power curves of t and the normal variate would coincide. In fact, it is interesting for the reader to make a direct comparison by superimposing the normal OC curves of Fig. 8-6 over those of the t test on Fig. 8-5. It becomes easy to observe in this connection that for the very small sample sizes of about four to seven the OC curves of the t test with n increased by two are about the same as those of the normal statistic! Then for somewhat larger sample sizes the OC curves of the t test for $(n + 1)$ nearly coincide with those of the normal statistic for sample size n. When the sample sizes get above $n = 20$ or more, the OC curves of t and the normal variate begin to coincide for the same sample sizes. This suggests that the normal approximation for sample sizes, i.e., Eq. 8-4 or Eq. 8-5, would be sufficiently accurate for many problems in the determination of sample sizes for t.

If we were to know sigma accurately and desire to test the hypothesis that the true mean of the normal population sampled is equal to a, i.e., $\mu = a$, versus an alternative that states that the true mean is not equal to a but rather that $\mu = \mu_1$, the sample sizes to control the error of rejecting the null hypothesis when true to the value α and the risk of not rejecting the null hypothesis when it is false and $\mu = \mu_1$ to the value of β are simply

$$n = [(z_\alpha + z_\beta)/\lambda]^2 \tag{8-52}$$

where λ is the departure of the true mean from a in standard deviations, i.e.,

$$\lambda = |\mu_1 - a|/\sigma. \tag{8-53}$$



Reprinted with permission from "Operating Characteristics for the Common Statistical Tests of Significance" by Charles D. Ferris, Frank E. Grubbs, and Chalmers L. Weaver, Annals of Mathematical Statistics XVII, No. 2 (June 1946). Copyright© by Institute of Mathematical Statistics.

Figure 8-6. Operating Characteristics of the Normal Test $z = \left[\frac{\sqrt{n}(\bar{x} - a)}{\sigma}\right]$ for Testing $\mu = a$ Against $\mu \neq a$ (Ref. 3)

Furthermore, if we were to sample two normal populations for the purpose of comparing their true mean values and they have a common known sigma, the sample size for this purpose is simply double that of Eq. 8-52 since the variance of the difference in sample means would be double that for a single mean. The critical region is always based on the significance level α chosen.

Now returning to Student's t test of Eq. 8-43 and our problem of controlling Type I and Type II errors to find the sample size, Neyman and Tokarska (Ref. 19) have contributed a key study of this problem and give samples sizes needed. On the other hand, Chand (Ref. 4) gives an approximate sample size, which is found from

$$n = [(z_\alpha + z_\beta)/\lambda]^2 + z_\alpha^2/2 + 1. \tag{8-54}$$

In a comparison of sample sizes based on Eq. 8-54 with those of Neyman and Tokarska (Ref. 19), Chand (Ref. 4) shows that the agreement is excellent even for the smaller sample sizes. However, the sample sizes based on the normal approximation of Eq. 8-52 are off about two for the smaller sample sizes of about four to seven. We note in this connection that Eq. 8-54 actually provides a correction to Eq. 8-52 in the form of adding half the normal significance level squared plus unity.

Although the straight normal approximation of Eq. 8-52 is off for the very smallest sample sizes, we might nevertheless consider the approximately normal Student's t with the divisor of $(n-3)$ for the SS deviations about the mean, that is, the quantity given by Eq. 4-105. The approximate normal variate is the z given in Eq. 8-46, so it should be clear to the reader that the actual sample size from this approximation is the normal sample size multiplied by the ratio $(n-1)/(n-3)$. This means, as the reader may check, that for the *smaller* sample sizes one adds about 2, i.e., n is given by

$$n \approx 2 + [(z_\alpha + z_\beta)/\lambda]^2 \tag{8-55}$$

and for sample sizes over about 25 we simply use the normal approximation of Eq. 8-52. Let us now give an example (Example 8-11).

Example 8-11:

Suppose that an acceptance test were being conducted for the 11 observed muzzle velocities for the 155-mm projectiles in Example 4-1. It was desired, furthermore, that the true mean velocity of the projectiles should be the nominal velocity of 2500 ft/s but no lower. Thus if the true or large sample muzzle velocity of the projectiles were, say, 2480 ft/s, one should have a very high assurance that the lot sampled for firing should be rejected. With these data find the sample size required in such a test to control risks of erroneous judgment to about 5% each.

Although we may have some idea concerning the size of the round-to-round standard deviation from the previous firing of Example 4-1, we should be cautious concerning the firing of only 11 rounds for either acceptance or rejection of an expensive lot of ammunition. Since the round-to-round standard deviation in muzzle velocity is expected to be about 10 ft/s, we will use this for sigma. Also we have that $\lambda = 2$. Using the straight normal approximation of Eq. 8-52, we get $n = 2.7$, which we know is too small, and hence we know that Eq. 8-55 would give 4.7, so we take $n = 5$. The reader may also note that Eq. 8-54 gives $n = 5.06$ (we use $n = 5$). Therefore, to our surprise, a sample of size five would meet our specified risk requirements. (One may note that the sigma of an average is $10/\sqrt{5} = 4.5$ ft/s, and we are picking up a deviation of over four times this value.)

8-6.2 SAMPLE SIZES FOR COMPARING THE MEANS OF TWO NORMAL POPULATIONS

After determining the sample size for making inferences about the single normal population mean, our purpose is to find sample sizes relating to the problem of comparing two normal population true means when the common standard deviation is unknown. Recall that significance tests for comparing normal population means were discussed in Chapter 4. This included the use of the F test to establish that the two normal

populations sampled had a common (or equal) variance(s), and it also included the Behrens-Fisher-type problem for the case in which the two unknown population variances might be unequal, as in par. 4-7.3.2. Student's t statistic for conducting a comparison of two means in a significance test for equal sigmas was discussed in par. 7-3.1. First, however, we will start with the comparison of two unknown normal population means for the case in which the variances are equal and accurately known. The notation for this particular case is as follows:

σ = known population sigma of the two populations
$\bar{x}_1$ = sample mean of first population sample
$\bar{x}_2$ = sample mean of second population sample
μ_1 = first population unknown true mean
μ_2 = second population unknown true mean
α = risk of rejecting the null hypothesis that $\mu_1 = \mu_2$ when true
β = risk of accepting the null hypothesis when actually $\mu_2 > \mu_1$ or H_1 is true
$\lambda = |\mu_2 - \mu_1|/\sigma$.

If it is assumed that the sample sizes are equal, i.e., $n_1 = n_2 = n$, the sample statistic used for testing the null hypothesis that the two true means are equal is simply

$$z = (\bar{x}_1 - \bar{x}_2)\sqrt{n}/\sqrt{2}\sigma. \tag{8-56}$$

The $\sqrt{2}$ in the denominator is necessary because we are dealing with the standard deviation of the difference between two sample means. Thus without going through the usual derivation, we can see immediately that the needed sample size is

$$n = 2[z_\alpha + z_\beta)/\lambda]^2. \tag{8-57}$$

Note that the sample size to control the stated risks is now double that for the single sample case of sampling only one normal population. Moreover, it also should be clear that when the variance is doubled, the sample size must be doubled also.

When the sample sizes and the two population sigmas are unequal, but the sigmas nevertheless are known accurately, the sample test statistic for equality of the two normal population true means is

$$z = (\bar{x}_1 - \bar{x}_2)/(\sigma_1^2/n_1 + \sigma_2^2/n_2)^{1/2}. \tag{8-58}$$

A solution may still be found if we know the k for which $\sigma_2 = k\sigma_1$ and the relation between the two n's, i.e., $n_1 = dn_2$. Moreover, Ferris, Grubbs, and Weaver (Ref. 3) point out that the Type II error β may easily be found from Fig. 8-6 for any λ'—say, n_1 and n_2, and k—by selecting the OC curve for any convenient sample size n and taking

$$\lambda = \lambda'(n_1 n_2)^{1/2}/[n(k^2 n_1 + n_2)]^{1/2}. \tag{8-59}$$

In summary, rather complete knowledge of the relation between the sigmas and the ratio of the n's must be known for this situation.

The more prevalent and important case for comparing two normal population true means concerns the situation for which we have no knowledge about either the relative size of the variances or the true means. We will, however, have established that the two normal populations have a common standard deviation or will resort otherwise to the Behrens-Fisher test of par. 4-7.3.2. For the case of equal sigmas or a common sigma, the determination of sample size is somewhat more complicated than for the case of known sigmas; and for the unequal sigma case requiring the Behrens-Fisher statistic, the best choice is probably to use the normal approximation for which the quantity $(n - 3)$ instead of the actual number of df equal to $(n - 1)$ is used as the divisor in Student's t statistic.

When it has been established on the basis of an F test that the two normal population sigmas are practically equal, one proceeds to calculate Student's t statistic

$$t = (\bar{x}_1 - \bar{x}_2)\sqrt{n}/(s\sqrt{2}) \tag{8-60}$$

where the quantity s^2 is the unbiased estimate of the common variance as in Eq. 4-108 and we assume equal sample sizes n. For this very prevalent case Chand (Ref. 4) suggests that the proper sample size to take from each of the two populations should be found from

$$n = \frac{b + [b^2 - 8\lambda^2/(z_\alpha + z_\beta)^2]^{1/2}}{2\lambda^2/(z_\alpha + z_\beta)^2} \tag{8-61}$$

where the value of b is

$$b = 2 + (1 + z_\alpha^2/4)\lambda^2/(z_\alpha + z_\beta)^2. \tag{8-62}$$

If one endures the algebraic trouble of substituting for b from Eq. 8-62 into Eq. 8-61 and simplifies as much as possible to two terms involving the expansion of the square root term, he will find that Eq. 8-61 is approximately equal to the normal approximation of Eq. 8-57 plus about 2!

To add to this enlightenment, one might well consider the Smith (Ref. 20) approximately normal statistic of Eq. 4-124 for comparing two unknown normal population means assuming no knowledge of the two sigmas—and which he will find in consonance with what we established previously—that the sample size may be taken as approximately equal to the numerical value determined from Eq. 8-57, which we further multiply by $(n-1)/(n-3)$ by using the n from Eq. 8-57. Thus we may now establish a rather general rule for the calculation of sample sizes. First, calculate n from Eq. 8-57 and use it if n exceeds about 20 or 25. Otherwise, and especially if n is perhaps 15 or less, multiply by $(n-1)/(n-3)$; or if you like, use the normal approximation of Eq. 8-57 and multiply by the quantity $(n-1)/(n-3)$ to obtain the final n! If n from Eq. 8-57 is very small, say four or five, then add two!

As a point of particular interest, the reader may have observed by now that the determination of sample size often seems to be detached from a given problem. For example, if one faces the problem of determining sample sizes for mean values, he very often must take the sum of the two upper probability levels of the standard normal distribution, divide this sum by the difference between the desired and undesired mean levels (which must be expressed in standard units), and then square the result for the single sample case. If he is dealing with the two-sample case, he merely doubles this answer! Of course, the problem of dealing with the ratio of two sigmas seems a bit different, but the normal approximations work very well there too! Example 8-12 illustrates the process of determining sample sizes for comparing the means of two normal populations.

Example 8-12:

Consider the data of Example 4-8 relative to a comparative test of current standard mechanical time fuzes used for reference purposes and a "better" fuze proposed by a manufacturer to replace the reference lot when exhausted. In this connection, it would seem that the mean value of 4.8 s for the proposed fuze is a bit low, and perhaps such a lot should be rejected, i.e., not used for reference purposes. The sample size of 10 would appear to be quite small and perhaps would give a flawed judgment! If we were to set the risks of erroneous judgment concerning the new lot of reference fuzes at, say, 2.5% and were to desire to pick up a difference of 0.10 s between mean values of the new lot of mechanical fuzes and the current reference lot (which has always been quite satisfactory), what sample size of each lot should we test?

To begin with, we should use whatever information can be gleaned from the data of the previous test. We note that the standard deviation of an individual fuze appears to be about 0.03 s less than that of the current reference lot although such an observed difference may not be significant. In any event, we have some reason to believe that a standard deviation of about 0.10 s should be quite satisfactory, and it will be difficult and perhaps costly to produce better fuzes. Thus we may as well take sigma equal to 0.10, and the difference of 0.10 s in which we are interested amounts to one sigma.

Next, if there were a difference in standard deviations of the current and proposed fuzes, perhaps an F test for large samples would show this, and perhaps the Behrens-Fisher test of par. 4-7.3.2 should be used. On the other hand, we know that the approximate test of Smith given by Eq. 4-124 and detailed in Ref. 20 takes care of different standard deviations very well. Therefore and in summary, we propose to use the normal approximation of Eq. 8-57, determine what it gives for n, and perhaps multiply the result by $(n-1)/(n-3)$. Finally, if we become a bit puzzled, we could calculate n from Eq. 8-61, which, however, assumes equal sigmas. If we need to get very fussy about the sample size, perhaps we need to do a bit of research to see whether one of the Behrens-Fisher types of tests would give a different—larger—sample size.

Since we are using the 97.5% probability levels, both standard normal deviates are equal to 1.960; by using Eq. 8-57 with $\lambda = 1$, we find that $n = 31.7$. If we multiply this value of n by $(n-1)/(n-3)$, i.e., $29.73/27.73$, we get the final $n = 32.95$, or 33. On the other hand, assuming equal sigmas (and we are somewhat assured they will be about 0.10 s), we find from Eq. 8-61 that $n = 31.7$; therefore, we conclude a sample size of about 31 would be quite appropriate.

This completes our coverage of the problem of sample size determination for the more common statistical tests of significance, which are carried out in many experimental situations. We believe that our presentation of this coverage will be useful for most of the applied problems the analyst will face in sampling a binomial or normal population. However, we will now devote a little attention to sample sizes and the power function or OC curves for the ANOVA test.

8-7 POWER FUNCTION AND SAMPLE SIZES FOR THE ANALYSIS OF VARIANCE TESTS

Although Student's t statistic is used to compare two unknown normal population means, the Snedecor-Fisher F test is used for the purpose of making judgments concerning whether or not several normal population means can be considered to be equal.

We will consider an ANOVA for samples of size n drawn from each of m normal populations, which are assumed to have the same variance, either for the observations on their original scale or after a variance-stabilizing transformation. The requirement is to decide, on the basis of the sample results, whether or not an undesirable amount of variation among the true means of the m normal populations exists. The usual test is the F test of significance—calculated by taking the SS of the m sample means about the grand average, converted to the equivalent variance of an individual observation, and divided by the number of df $(m-1)$; this result is divided by the SS within the m samples divided by the $m(n-1)$ df. Thus we define

x_{ij} = ith item (observed value), $i = 1,2,\ldots,n$, of the jth sample, $j = 1,2,\ldots,m$, drawn at random from the jth normal population

$\bar{x}_{.j} = \sum_{i=1}^{m} x_{ij}/n$ = sample average from the jth population

$\bar{x}_{..}$ = grand average of all mn observations.

The calculated F statistic or ratio is

$$F = nm(n-1)\sum_{j=1}^{m}(\bar{x}_{.j} - \bar{x}_{..})^2 \Big/ \left[(m-1)\sum_{i=1}^{n}\sum_{j=1}^{m}(x_{ij} - \bar{x}_{.j})^2\right]. \tag{8-63}$$

Thus we are dealing, for illustrative purposes, with a one-way classification in the ANOVA although the use of the F ratio here could be considered to be much more general. If the observed F in Eq. 8-63 exceeds the (upper) significance level chosen, we conclude that the population means are not equal. If some of the normal population means are unequal, there is an additional component of variance among them as contrasted to the residual or "within" variance σ^2 of the m normal populations sampled. Therefore, we might say that if the null hypothesis of no difference in levels of the m normal populations is invalid, we may describe this additional variance as, say, $\theta^2\sigma^2$. Then under the null hypothesis we have

$$H_0 : \theta = 0.$$

And under the existence of an invalid H_0, so that the alternative hypothesis of unequal true means prevails, we see that

$$H_1 : \theta > 0.$$

Moreover, since the variance among sample means of n observations is σ^2/n, the total variance among the sample means if H_1 holds is

$$\sigma^2/n + \theta^2\sigma^2 = \lambda^2\sigma^2/n \tag{8-64}$$

where we have set

$$\lambda^2 = 1 + n\theta^2. \tag{8-65}$$

At this point, we must consider that there are two possible models for the ANOVA. First, there is the Model I, or fixed-effects model, for which our interest is only in the particular m treatments we tested in the experiment. Therefore, for the fixed-effects model we might, for example, be interested in which of the m treatments is the superior one and not regard a hypothesis that covers the possibility that m treatments may have been random selections from a larger population or universe. There is also the Model II, or random-effects model, for which we assume that the m treatments are chosen at random from a universe of their own, so that the sample results may be used to infer characteristics of the "population of means" from which the m samples were randomly drawn.

For the Model I, or fixed-effects case, the power function of the ANOVA test has been thoroughly studied by Tang (Ref. 21), who tabulated the relevant characteristics of it. Thus we refer interested readers to that publication for dealing with the case of a (relatively small) number of fixed effects.

For our limited purposes we will cover only some points concerning the random-effects means, i.e., Model II. For this case, the quantity F/λ^2 follows the F distribution with $(m-1)$ and $m(n-1)$ df, respectively, so that the OC curves of F could be entered to find, for various values of θ, the chance of accepting the null hypothesis H_0 when the alternatives H_1 are true. Thus such OC curves for F would have to be for different sample sizes, $n_1 = m$ and $n_2 = nm - m + 1$, with the value of λ for entering the curves given by Eq. 8-65. The OC curves for F on Fig. 8-4 are for equal sample sizes, whereas for illustrative purposes we indicate on Fig. 8-7 just how the OC curves may appear for certain cases of the sample sizes. Additional OC curves for some other sample sizes are given in Refs. 1 and 3.

As an example of the use of such OC curves as those depicted for certain sample sizes on Fig. 8-7, suppose we were faced with the design of some experiment for which the number of populations to be sampled were rather indefinite and the total sample size of the experiment had to be limited to, say, $mn = 24$. Then our division of the $mn = 24$ into the number m of samples and the size n of each would depend on the size of θ we would like to be very positive of detecting if H_1 were actually true. This particular computation has been made by Ferris, Grubbs, and Weaver (Ref 3), and we give their informative table here as Table 8-3 for the best division of only 24 test observations. Note that for the smaller values of θ one places more emphasis on the estimation of sigma by sampling only two or three of the possible different normal populations. On the other hand, and as the value of θ becomes large, i.e., there is quite a significant variation among the true means of the populations, more emphasis should be directed toward sampling as many of the different populations as possible.

Perhaps there will be a large number of cases for which the approach and tables of Tang (Ref. 21) will be required; on the other hand, the rather simple Model II approach covered here also may be found useful perhaps as a preliminary calculation to more or less decide on the particular Model I experiment. Moreover, the random treatments dealt with here for Model II obviate the difficulties imposed by the noncentral chi-square distribution, and there often will be cases in which one will want to know the relative sizes of components of variance in many experimental situations. For a suitably large number of treatments, it can be said that the Model I case approaches that of Model II.

TABLE 8-3

VALUES OF m, n, AND θ FOR BEST POWER WHEN $mn = 24$ (Ref. 3)

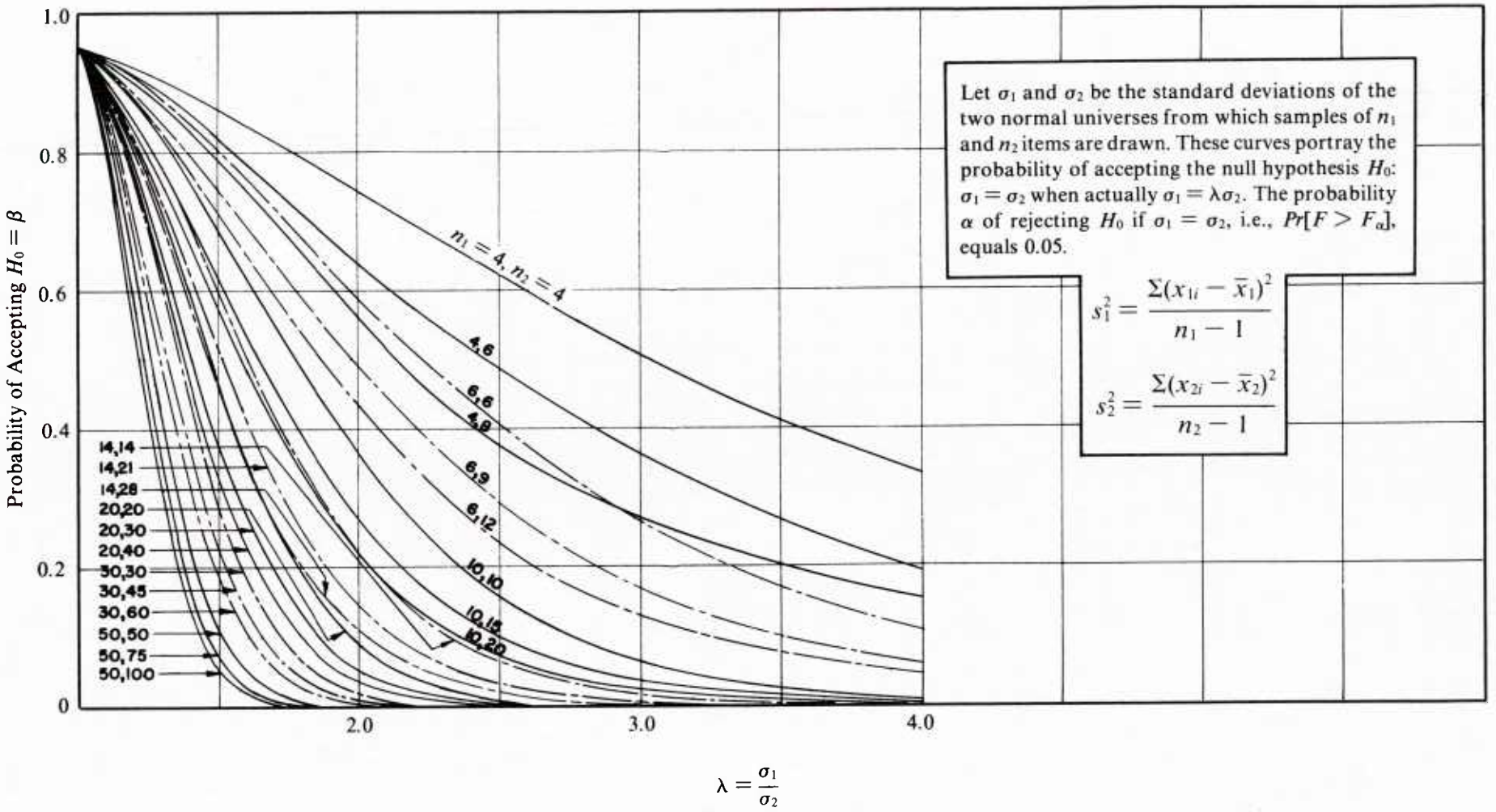
m	n	θ
2	12	0.00 - 0.32
3	8	0.32 - 0.60
4	6	0.60 - 0.91
6	4	0.91 - 1.37
8	3	1.37 - 2.50
12	2	2.50 -

Reprinted with permission from "Operating Characteristics for the Common Statistical Tests of Significance" by Charles D. Ferris, Frank E. Grubbs, and Chalmers L. Weaver, Annals of Mathematical Statistics XVII, No. 2 (June 1946). Copyright©by Institute of Mathematical Statistics.

For numerous practical or experimental situations, one does not have to know either the power of the test or the sample size exactly; good approximations will be sufficient. In this connection, Pearson and Hartley (Ref. 22) have provided charts or graphs of the power function, derived from the noncentral F distribution, for the ANOVA technique, and these should be adequate for most experimental situations. The reader is urged to use these charts, at least as a first try.

Guenther (Ref. 23) has made a study of the power and sample size determination problem when the alternative hypotheses are given in terms of the quantiles of normal distributions. In fact, the power of normal-theory tests about mean values of populations depends on a noncentrality parameter, which unfortunately is a function of the unknown parameter sigma. Hence to calculate the power and solve sample-size problems, one usually expresses differences in mean values in terms of the unknown sigma, which overcomes this problem and is quite natural anyway since sigma characteristically is the parameter that well describes the width of the distribution sampled. Guenther (Ref. 23) points out that one may express alternative hypotheses in terms of quantiles. In other words, instead of hypothesizing that the mean of one normal distribution is greater than that of another, one could say that the 50% point of one normal distribution is at the same level as the 60% point of another normal distribution, which is another way of describing that the mean of the first normal population exceeds that of the second one. Furthermore, Guenther points out in his paper that the quantile approach also eliminates the unknown population sigma from the problem. In Ref. 23 Guenther covers the problem of sample-size determination using quantiles for sampling a single normal population, comparing two normal population means, or making hypotheses about the true means in a one-way classification of the ANOVA. He also covers several treatment means for randomized complete blocks. The key parameter used for the alternative hypotheses is expressed in terms of the SS of deviations of the true means from a central mean. Thus the reader may also want to consider this approach for the determination of sample sizes in the more complex experiments or even for some of the common statistical tests of significance.

Odeh and Fox (Ref. 24) have published a series of charts for dealing with the sample-size-choice problem for tests of statistical hypotheses in connection with designing experiments. As we have indicated, one should consider both the significance level and also the power (or OC curve) of the test for experimental comparisons. One can control both of these quantities by selection of the number n of replicates since the power for a fixed significance level α increases as the sample size n increases. The Odeh and Fox charts of Ref. 24 are designed to enable one to find the proper sample size n for a given α and desired power in experiments for which linear models are appropriate. A wide range of both significance levels and degrees of power are covered in Ref. 24. Ref. 24 also has extensive tables of the percentage points of both the F and chi-square distributions, and in addition the tables give pertinent references concerning previous tables, charts and programs, and many examples and computational methods.



Reprinted with permission from "Operating Characteristics for the Common Statistical Tests of Significance" by Charles D. Ferris, Frank E. Grubbs, and Chalmers L. Weaver, Annals of Mathematical Statistics XVII, No. 2 (June 1946). Copyright © by Institute of Mathematical Statistics.

Figure 8-7. Operating Characteristics of the F-Test $\left[F = \frac{s_1^2}{s_2^2}\right]$ for Testing $\sigma_1 = \sigma_2$ Against $\sigma_1 > \sigma_2$ ($n_1 = n_2$, $2n_1 = 3n_2$, $n_1 = 2n_2$) (Ref. 3)

8-8 SOME ADDITIONAL DISCUSSION ON SAMPLE SIZE DETERMINATION FOR ATTRIBUTE AND NORMAL POPULATION SAMPLINGS

Although we have covered a considerable amount of the statistical literature concerning sample size determination for binomial and normal population sampling in connection with some typical Army experiments, there are some additional topics the Army analyst might find of interest in his work.

With reference to an extension of binomial-type comparisons to the analysis of multinomial populations, Guenther (Ref. 25) gives some very pertinent discussions relevant to the power and sample size for approximate chi-square tests that are so often used in multinomial comparisons. In Ref. 25 Guenther presents methods of power calculation and sample-size determination and then illustrates the three most frequently used types of the multinomial comparison tests. These include the specification of multinomial p's under the alternative hypothesis, the test of independence in association, and the test of homogeneity. Such calculations involve noncentrality parameters of chi-square, as would be expected, and these are given in Guenther's paper (Ref. 25) together with an example of each of the three types of multinomial analyses. In particular, the tables of the cumulative noncentral chi-square of Haynam, Govindarajulu, and Leone (Ref. 26) are found to be very useful.

For the binomial type of sampling inspection, Hahn (Ref. 27) discusses the problem concerning what is the smallest number of units that need be sampled from a lot for the probability to be at least a given percent that the lot will be rejected if it contains p percent or more defectives. These particular sampling plans call for zero observed defectives in the sample for the lot to pass inspection. Thus an important shortcoming of the "minimum size sampling plans" is that the percent defective of the lot sampled must be appreciably lower than the allowed percent defective for there to be a high probability of passing the inspected lot. The reader may check this by calculating the OC curve for any zero defects single sampling inspection plan, and he should note that the OC curve comes down very sharply for increasing values of the percent defective in the lot.

The subject of order statistics and the many types of applications to various Army problems are discussed in Chapter 7. One of the rather important topics presented in par. 7-7.5 is tolerance intervals. Recall that the tolerance interval is that interval of the largest and smallest sample values, for example, and no matter what the distributional shape, one can make a confidence statement that the tolerance interval includes a certain percentage of the distribution sampled. Eq. 7-31 gives the relation between the sample size, the proportion of the population covered by the tolerance interval, and the confidence level stated or desired. Hence it is seen that Eq. 7-31 may be used to determine the sample size necessary to include a desired percentage of the population sampled for a given level of confidence. We record here that the determination of the sample size may be by the methods of Guenther in Ref. 10 or in Ref. 28. There would seem to be many Army applications for which such sample sizes are desired.

If one is interested in the determination of sample sizes for tolerance intervals on the normal population sampled, it is suggested that he study Faulkenberry and Daly's paper (Ref. 29). They discuss both the one-sided and two-sided types of tolerance intervals. If one knows that the sampled population is normal, this would lead to either a shorter tolerance interval for the same sample size as that used to sample a general unknown distribution, or for the same width of tolerance interval, the sample size would be smaller for the known normal distribution than for a general unknown shape. Thus this represents another area of application for which sample sizes are important.

Returning to the sampling of two normal populations to make a comparison of their mean levels, or especially to determine whether one of the normal populations generally exceeds the other in level of operation, Guenther (Ref. 30) discusses the determination of sample sizes when one desires to make the comparison on the basis of quantiles. As stated in Ref. 30, Guenther shows that the solution to this problem depends on the noncentral t distribution, and he establishes a rather simple equation (his Eqs. 2.7 and 2.9) for the estimation of sample size. An instructive example with a detailed solution is also given by Guenther in Ref. 30.

As contrasted with the ordinary ANOVA technique using the F test to judge whether the means of several normal populations are equal, Bechhofer, for example, in Ref. 31, began a series of studies relative to multiple decision procedures for ranking the means of normal populations that are sampled for the purpose. In this connection, one may sample to some predetermined sample size and then stop to make a judgment concerning the ordering of the normal population means. A number of papers have been published on this procedure for

sampling several normal populations, and in fact, the amount of literature has grown rather extensively. For the sample size determination problem we suggest that interested readers consult the paper of Ramberg (Ref. 32). Ramberg gives two conservative sample-size approximations for this particular sampling procedure. Although there could be many applications of the Bechhofer type of sampling to rank normal population means and/or variances, we cannot explore this area of investigation any more thoroughly.

In addition to our somewhat useful—but also rather incomplete—account of the sample size determination problem in general, the reader will have noticed that we limited our discussion to some of the more usual types of statistical problems he will face in day-to-day work. However, we also should remark that many other problems exist that require the *a priori* selection of sample size in order to conduct an experiment properly. For example, in addition to our coverage there is the whole area of curve fitting, least squares, and regression applications. The determination of sample sizes for these types of problems has not appeared very widely in the statistical literature as yet although it is expected that more and more papers on this and other subjects will appear. An example of a study on the selection of sample size for regression analysis is that of Park and Dudycha (Ref. 33). Park and Dudycha (Ref. 33) have developed what they refer to as a "cross-validation" approach to determine sample sizes for regression models. They discuss both the fixed model case, for which it is assumed that the independent variables are (mathematical) quantities free of error, so to speak, and also the random model, which refers to the case for which the dependent variable y is predicted in terms of random variables x_i, which follow the multivariate normal distribution. Several tables are given in Ref. 33 to aid in the selection of sample sizes. This type of problem can become rather involved when one also may have to select several variables from many possible independent ones in the course of his regression studies.

So far, our sample size determination discussion has centered around two of the more important distributions in much analytical work, i.e., the binomial and the normal distributions. Nevertheless, we think it desirable to include some limited account of sample sizes for sampling exponential distributions. In this connection, there are many problems in the currently important fields of reliability and life testing that also require selection of sample sizes. Therefore, it seems advisable to give some guidance in these areas before completing this chapter on sample sizes.

8-9 SAMPLE SIZES FOR EXPONENTIAL POPULATIONS

Many Army applications of statistical methods are related to the sampling of exponential populations especially in the areas of reliability analyses and life testing situations. Hence there are occasions for which the determination of appropriate sample sizes is of interest either for estimation purposes or for controlling Type I and Type II errors. Since lifetimes are generally taken on the basis of a time scale, we will use t as the measured random variable. Thus the exponential probability density function (pdf) of lifetimes is taken as

$$f(t) = (1/\theta) \exp(-t/\theta) \tag{8-66}$$

where

θ = true unknown time-to-fail for the items tested.

Moreover, if one were to put n items following the exponential distribution on test and measured the lifetimes of the first r failures, at which point the test is truncated, it is well-known that the minimum variance, best unbiased estimator $\hat{\theta}$ of the parameter θ is

$$\hat{\theta} = [\sum_{i=1}^{r} t_i + (n-r)t_r]/r. \tag{8-67}$$

Moreover, the quantity

$$2r\hat{\theta}/\theta = \chi^2(2r) \tag{8-68}$$

follows the chi-square distribution with $2r$ df. Thus with the aid of Eq. 8-68 one can place a confidence bound about the unknown parameter θ and hence obtain the sample size or, more appropriately in this case, the

number of failure times r necessary to estimate θ within any desired limits. For example, suppose we wanted to determine the number of failures r such that the estimate of Eq. 8-67 will be within $q\theta$ of the true unknown θ, where $q = 1\%, 5\%$, etc. Then this would mean that the bounds on $\hat{\theta}/\theta$ would be $(1-q)$ to $(1+q)$ so that we could equate the difference between the upper and lower $\alpha/2$ probability levels of chi-square to the difference $[2r(1+q) - 2r(1-q) = 4qr]$ to obtain the needed number of failures r to achieve the estimate $\hat{\theta}$ of θ. The reader may verify that

$$r = z_{1-\alpha/2}/(4q) \tag{8-69}$$

where we have used Fisher's square root transformation of chi-square to approximate normality. Thus for having $(1-\alpha)$ confidence of getting the estimate Eq. 8-67 within a fraction q of the unknown true parameter of the exponential distribution, one must sample until the number of failures is equal to the upper $\alpha/2$ level of the standard normal distribution divided by four times the quantity q. Example 8-13 is helpful at this point.

Example 8-13:

Past experience indicates that the number of miles to failure for an M111 personnel carrier is believed to follow an exponential distribution. It is desirable in this connection to know within 5% just what is the mean number of miles to failure for the population of M111 vehicles. Therefore, determine the number of failures that must be observed to establish the mean-miles-to-failure within 5%.

For this problem, let us decide to use the 95% level of confidence and the two-sided test, i.e., we merely want the estimate to deviate either above or below the true value by no more than 5%. Thus we see that $+z_{\alpha/2} = 1.96$, and from Eq. 8-69

$$r = 1.96/(4 \times 0.05) = 9.8, \text{ or use } r = 10 \text{ failures.}$$

Now recall that we are dealing with the number of failures required and not the sample size, which may be greater. Hence we could put only 10 vehicles on test and run them until all 10 have failed and estimate the parameter θ from 10 failure times, using Eq. 8-67 with $r = n$. Better still, to save time, put about $n = 15$ or more vehicles on test until the number of observed failures is $r = 10$ and stop the test, using Eq. 8-67 with $n = 15$ (or whatever) and $r = 10$.

It might be interesting to point out for this example that one could logically be interested only in being 95% confident that the true value of the parameter of the exponential distribution will not fall below the estimate by more than 5%. In this case the value of r would be

$$r = 1.645/(4 \times 0.05) = 8.3 \text{ failures.}$$

Thus if we ran the test until eight failures occurred, we would have less than 95% confidence that the true θ would not be below the estimate by more than 5%, and if we were to continue the test until nine failures occurred, then our confidence would exceed 95%.

Now we will discuss the determination of the required number of failures for guarding against a low error of rejecting the null hypothesis when true and also a low error of accepting the null hypothesis when it is false and an undesirable value of the unknown parameter prevails.

We will refer to the acceptable value of the mean life under the null hypothesis as θ_0 and have

$$H_0 : \theta = \theta_0.$$

On the other hand, for the alternative hypothesis

$$H_1 : \theta = \theta_1$$

and since we will usually desire that the mean life of an item, component, system, etc., be as long as possible, it becomes important to guard against the possibility that the true unknown mean life θ is as low as the undesirable θ_1 $(\theta_1 < \theta_0)$.

It is well-known from the exponential life testing theory of Epstein and Sobel (Ref. 34) that the power function relation between the parameters and the number of failures depends on chi-square and is given by

$$\theta_0/\theta_1 = \chi^2_{1-\beta}(2r)/\chi^2_{\alpha}(2r) \tag{8-70}$$

where $\chi^2_{\alpha}(2r)$ is the lower α probability level of chi-square and $\chi^2_{1-\beta}(2r)$ is the upper β probability level with $2r$ df each. Thus we see that the problem of determining the sample size for the exponential distribution is very similar to that encountered elsewhere in this chapter, as for the comparison of variances from a normal population. As before, a number of suitable approximations to chi-square may be used to estimate the needed sample size for the desired protection. In Ref. 35 the Wilson-Hilferty transformation of chi-square to an approximate normal variable was used in the interest of rather accurate calculation of probabilities. (The Wilson-Hilferty, or cube root, transformation of chi-square is covered in their paper, Ref. 36.) However, for the calculations of the required number of failures, the more accurate Wilson-Hilferty transformation is unnecessarily complicated, and some simpler approximations are quite satisfactory for sample size or number of failures computations. We shall present them.

Let us take

$$\lambda = \theta_0/\theta_1 \tag{8-71}$$

that is, λ is the ratio of the desired mean life θ_0 to the undesirable value θ_1 of the mean life. Then for the defined quantities

$$\delta = (\theta_0/\theta_1)^{1/3} \tag{8-72}$$

$$\eta = (z_{1-\beta} + \delta z_{\alpha})/(\delta - 1) \tag{8-73}$$

Grubbs (Ref. 35) shows that the number of failures to control errors to α and β, respectively, may be found from

$$r = (4/9)[(\eta^2 + 4)^{1/2} - \eta]^2. \tag{8-74}$$

Narula and Li (Ref. 37) have investigated a number of simpler approximations for this particular problem and have found that all of the approximations are close together, especially when the calculated r's are rounded upward to the next integer. Hence, for example, one may as well use the simpler normal approximation given by

$$r = [(z_{1-\beta} + \theta_0 z_{\alpha}/\theta_1)/(\theta_0/\theta_1 - 1)]^2. \tag{8-75}$$

For an example (Example 8-14), we will use the same one as in Ref. 35 for the mean-miles-between-failures (MMBF) of some "main battle tanks". Although Ref. 35 used the rather complex approximation of Eq. 8-74, we will use the simpler normal approximation of Eq. 8-75.

Example 8-14:

Suppose we would like to test some main battle tanks to determine whether as a class they have a MMBF of 600 mi or a MMBF of only 300 mi. We set a risk of 5% of rejecting the null hypothesis MMBF = 600 mi when true and a risk of 10% of accepting the null hypothesis MMBF = 600 mi when actually the true unknown MMBF is only 300 mi.

With this statement of the problem, our basic data are

$$\alpha = 0.05 \qquad \beta = 0.10$$
$$\theta_0 = 600 \qquad \theta_1 = 300.$$

Hence the required number of failures is calculated from Eq. 8-75 as

$$r = \{[1.282 + (2)\,(1.645)]\,/\,(2 - 1)\}^2 = 20.9$$

which is about 1 unit larger than the more "accurate" number (20) of failures to observe computed from Eq. 8-74. However, Narula and Li (Ref. 37) also recommend an improved approximation over that of Eq. 8-75, which is

$$r = \{(z_{1-\beta} + z_{1-\alpha})/[\ln(\theta_0/\theta_1)]\}^2 \qquad (8\text{-}76)$$

which for our data gives a value of $r = 17.84$, and this number rounded up to 18 gives a number of failures that is two less than the corresponding value from the more exact number of failures calculated from Eq. 8-74.

The reader should not regard any of these approximations as "exact", and one should expect that differences of this order might occur. In fact, the normal approximations are simple indeed, but it cannot be expected that they are exact in any sense. We believe that the use of the Wilson-Hilferty transformation of chi-square to an approximate normal variate, which is used in Eq. 8-74, should generally be more accurate because it has been checked widely, especially insofar as probabilities are concerned. However, it is more complex than the other approximations, and on practical grounds one might argue that it is not worth the extra effort for a difference of one or two units.

8-10 SUMMARY

The determination of sample sizes in Army problems is a very important and always timely problem because the aim is usually to save on the amount of testing and the number of dollars expended. In day-to-day applications the analyst faces many problems that involve the determination of sample size for the more common statistical tests of significance. However, in the future there will be more and more requirements to estimate sample sizes needed for the more complex types of experiments. Therefore, we have endeavored in this chapter to give a good introduction to both areas.

Two primary methods for the determination of sample sizes were discussed—the first was sample size selection on the basis of a high level of confidence that a given difference of departure from our expectation will be detected, and the other method considers the technique of controlling errors of judgment. This latter procedure has as its aim the setting of allowable errors—along with preselected confidence levels—for rejecting the null hypothesis when it is true and the acceptance of the null hypothesis when it is actually false and an alternative is true. This approach, it seems to us, is the more sound one to select for many important Army problems. The sample sizes for low risks, low rates of prematures, or high reliability may in some cases be prohibitive; accordingly, engineering judgment often must be applied along with the statistical considerations.

Insofar as possible, we have endeavored to record the simpler equations for sample size determinations and take into account the need for quickness in making the required calculations. In many cases the sample size required is a calculation calling for the ratio of the sum of two normal percentage levels or points in the numerator divided by the appropriate standard deviation of the difference in two estimated values. Often, such a calculation of the sample size may not be off more than a unit or two.

We realize that much additional research will be needed in connection with the determination of sample sizes for all types of Army applications. Hopefully, this account will not only give an introduction to the problem but also will serve to stimulate much additional thinking on this ever-important subject.

REFERENCES

1. AMCP 706-110, Engineering Design Handbook, *Experimental Statistics, Section 1, Basic Concepts and Analysis of Measurement Data,* December 1969.
2. AMCP 706-111, Engineering Design Handbook, *Experimental Statistics, Section 2, Analysis of Enumerative and Classificatory Data,* December 1969.
3. Charles D. Ferris, Frank E. Grubbs, and Chalmers L. Weaver, "Operating Characteristics for the Common Statistical Tests of Significance", The Annals of Mathematical Statistics **17**, 178-97 (June 1946).
4. Uttam Chand, "On the Derivation and Accuracy of Certain Formulas for Sample Sizes and Operating Characteristics of Nonsequential Sampling Procedures", Journal of the National Bureau of Standards **47** (Research Paper 2277), 491-501 (December 1951).
5. *Standard Recommended Practice for Choice of Sample Size to Estimate the Average Quality of a Lot or Process*, ASTM Designation E 122-72, American Society for Testing and Materials, 1916 Race Street, Philadelphia, PA 19103.
6. Frank E. Grubbs, "On Designing Single Sampling Inspection Plans", The Annals of Mathematical Statistics **20**, 242-56 (June 1949).
7. Robert E. Clark, "Percentage Points of the Incomplete Beta Function", Journal of the American Statistical Association **48**, 831-43 (December 1953).
8. Gerald J. Hahn, Random Samplings, "How Large a Sample Do I Need For 95% Confidence?", Chemical Technology **5**, 61-2 (January 1975).
9. Gerald J. Hahn, "Minimum Size Sampling Plans", Journal of Quality Technology **6** (July 1974).
10. William C. Guenther, "Sample Size Formulas for Some Binomial Type Problems", Technometrics **16**, 465-7 (August 1974).
11. E. S. Pearson and H. O. Hartley, Editors, *Biometrika Tables for Statisticians,* Vol. 1, Cambridge University Press, Cambridge, England, March, 1954.
12. Leslie E. Simon, "The Relation of Engineering to Very High Reliability", *Proceedings of the Tenth National Symposium on Reliability and Quality Control.* Sponsored by IEEE, ASQC, and the IES, Washington, DC, 7, 8, and 9 January 1964.
13. DARCOM-P 706-101, Engineering Design Handbook, *Army Weapon Systems Analysis, Part I,* November 1977.
14. W. A. Thompson and Jack Andriss, "The Required Sample Size When Estimating Variances", The American Statistician **15**, 22-3 (June 1961).
15. M. S. Bartlett and D. G. Kendall, "The Statistical Analysis of Variance Heterogeneity and the Logarithmic Transformation", Supplement to the Journal of the Royal Statistical Society **8**, 128-38 (1946).
16. Bert van Uitert, "Sample Size for Chi-square Tests", The American Statistician **18**, 42 (October 1964).
17. Harry E. Allison, "Computational Forms for Chi-square", The American Statistician **18**, 17-8 (February 1964).
18. George J. Resnikoff and Gerald J. Lieberman, *Tables of the Noncentral t-Distribution,* Stanford University Press, Stanford, CA, 1957.
19. Jerzy Neyman and B. Tokarska, "Errors of the Second Kind in Testing 'Student's' Hypothesis", Journal of the American Statistical Association **31**, 318-26 (1936).
20. T. M. F. Smith, "Defining Student's *t*", Letter to the Editor of the American Statistical Association **26**, 43-4 (December 1972).
21. P. C. Tang, "The Power Function of the Analysis of Variance Tests With Tables and Illustrations of Their Use", Statistical Research Memoirs, University of London, **2**, 126-49 (1938).
22. E. S. Pearson and H. O. Hartley, "Charts of the Power Function for Analysis of Variance Tests, Derived from the Noncentral *F* Distribution", Biometrika **38**, 112-30 (1951).
23. William C. Guenther, "Power and Sample Size Determination When the Alternative Hypotheses Are Given in Quantiles", The American Statistician **31**, 117-8 (August 1977).
24. Robert E. Odeh and Martin Fox, *Sample Size Choice, Charts for Experiments With Linear Models,* Marcel Dekker, Inc., New York, NY, 1975.
25. William C. Guenther, "Power and Sample Size for Approximate Chi-Square Tests", The American Statistician **31**, 83-5 (May 1977).

26. G. E. Haynam, Z. Govindarajulu, and F. C. Leone, "Tables of the Cumulative Non-Central Chi-Square Distribution", Selected Tables in Mathematical Statistics **1**, Eds. H. L. Harter and D. B. Owen. Markham Publishing Co., Chicago, IL, 1970.
27. Gerald J. Hahn, "Minimum Size Sampling Plans", Journal of Quality Technology **6**, 121-7 (July 1974).
28. William C. Guenther, "Determination of Sample Size for Distribution-Free Tolerance Limits", The American Statistician **23**, 44-6 (February 1970).
29. G. David Faulkenberry and James C. Daly, "Sample Size for Tolerance Limits on a Normal Distribution", Technometrics **12**, 813-21 (November 1970).
30. William C. Guenther, "A Sample Size Formula for a Non-Central *t* Test", The American Statistician **29**, 120-1 (August 1975).
31. Robert E. Bechhofer, "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations With Known Variances", The Annals of Mathematical Statistics **25**, 16-39 (1954).
32. John S. Ramberg, "Selection Sample Size Approximations", The Annals of Mathematical Statistics **43**, 1977-80 (1972).
33. Colin N. Park and Arthur L. Dudycha, "A Cross-Validation Approach to Sample Size Determination for Regression Models", Journal of the American Statistical Association **69**, 214-8 (March 1974).
34. B. Epstein and M. Sobel, "Life Testing", Journal of the American Statistical Association **48**, 486-502 (1973).
35. Frank E. Grubbs, "Determination of the Number of Failures to Control Risks of Erroneous Judgments in Exponential Life Testing", Technometrics **15**, 191-3 (1973).
36. E. B. Wilson and M. M. Hilferty, "The Distribution of Chi-Squares", Proceedings of the National Academy of Sciences **17**, 684-8 (1931).
37. S. C. Narula and F. S. Li, "Sample Size Calculations in Exponential Testing", Technometrics **17**, 229-31 (May 1975).

BIBLIOGRAPHY

Claude Angers, "A Graphical Method to Evaluate Sample Sizes for the Multinomial Distribution", Technometrics **16**, 469-71 (August 1974).

Z. Asaki, Y. Kondo, and T. Narita, "On the Power of *F*-Test in Analysis of Variance - I; Power of *F*-Test on the Main Effects", Reports of Statistical Applied Research, Union of Japanese Scientists and Engineers **17**, 57-66.

T. L. Bratcher, M. A. Moran, and W. I. Zimmer, "Tables of Sample Sizes in the Analysis of Variance", Journal of Quality Technology **2**, 156-64 (1970).

Martin Fox, "Charts of the Power of the *F*-Test", The Annals of Mathematical Statistics **27**, 484-97 (1956).

M. A. Kastenbaum, D. G. Hoel, and K. O. Bowman, "Sample Size Requirements: Randomized Block Designs", Biometrika **57**, 573-7 (December 1970).

J. E. Mann, "Monotonicity of a Power Function: Binomial Case", The American Statistician **32**, 76-7 (May 1978).

Edwin G. Olds, "Power Characteristics of Control Charts", Industrial Quality Control **18**, 4-10 (1961).

D. B. Owen, *Handbook of Statistical Tables,* Addison-Wesley Publishing Co., Inc., Reading, MA, 1962.

M. L. Tiku, "More Tables of the Power of the *F*-Test", Journal of the American Statistical Association **67**, 709-10 (1972).

CHAPTER 9

SENSITIVITY ANALYSES OF QUANTAL RESPONSE TYPE DATA

The term "sensitivity analysis" has been used in recent years to describe the statistical analysis of quantal response (all or nothing) type data. There are many important Army problems, for example, the ballistic limit of armor plate or the sensitivity of explosives to impact, which require this type of analysis. Moreover, it becomes highly desirable to estimate the low or high percentage points of an underlying distribution of the proportions of responses so that efficient test strategies to estimate these percentage points, or even the parameters of the distributions, may become very important.

Likely underlying distributions of the normal, logistic, and Weibull models are treated analytically, and the more efficient methods of estimation are covered. The applicable theory should suffice for many of the sensitivity analysis problems that the analyst may face in practice. We present a variety of examples to illustrate just how sensitivity analysis theory developed over the years applies to typical problems.

9-0 LIST OF SYMBOLS

A = lower boundary of Langlie's test strategy
A = arbitrary point
a_i = series of constants in the Robbins-Monro approximation method (see Eq. 9-9)
a_i = substitution for x_i for which there are positive responses or penetrations
a_i, b_j, s_i, t_j = coefficients, constants, or transformations used by DiDonato and Jarnagin
B = upper boundary of Langlie's test strategy
b_j = substitution for x_j, for which there are nonresponses or nonpenetrations
c = constant
d = interval of interest
$E(x)$ = expected value of x
$E(\delta_i) = p_i$ = expected value of δ_i
$F = F(x)$ = cumulative probability distribution
$f(x)$ = probability density function of a random variable
i = denotes the ith trial
L = natural logarithm of likelihood function
L_p = designation of Einbinder to specify a percentage point
L_α = first partial derivative of logarithm of likelihood with respect to α
$L_{\alpha\beta}$ = second partial derivative of L with respect to both α and β
$L_{\sigma\sigma}$ = second partial derivative of L with respect to σ
$\text{logit}p_i = \ln(p_i/q_i)$ = denotes the logit transformation
m = number of nonpenetrations (see Eq. 9-14)
$N = n + m$ = sum of penetrations and nonpenetrations (see Eq. 9-14)
n = number of penetrations (see Eq. 9-14)
n = denotes sample size, number of items, number of levels
n_i = number of items tested at stress level x_i
n_0 = Wetherill's designation for the number of positive responses at a stress level before a change of stress (also a key parameter in Einbinder's test strategy)

n_1 = number of test specimens responding at x_1

n_2 = number of test specimens responding at x_2

OSTR = one-shot transformed response

P = transformation of Einbinder, such as p^2, p^3, etc.

$P = f(p)$ = function of p

$p = p(x) = F(x)$ = cumulative probability distribution or proportion

p' = alternative value of p

$p(s_i)$ = transformed p_i used by DiDonato and Jarnagin

$p_i = p(x_i) = F(x_i)$ = proportions of responses at stress x_i

p_1, p_2 = proportions of response corresponding to stress levels x_1, x_2

$q = 1 - p$, or a percentage

$q(t_j)$ = transformed q_i used by DiDonato and Jarnagin

r_i = observed number of responses at stress level x_i

r_1 = number of test specimens that respond at x_1

r_2 = number of test specimens that respond at x_2

$s_i = a_i\beta - \alpha$ = transformation used by DiDonato and Jarnagin

TMP = transformed median percentage

$t_i = b_j\beta - \alpha$ = transformation used by DiDonato and Jarnagin

$u_i = u_i(s_i)$ = designation of DiDonato and Jarnagin for a normal probability distribution function

u_i, v_j = designations of DiDonato and Jarnagin for normal probability density function (pdf)

Var( ) = denotes variance of the quantity in ( )

V_{50} or $V_{0.50}$ = striking velocity at which 50% of the projectiles penetrate the armor plate

$v_i = v_i(t_i)$ = designation of DiDonato and Jarnagin for normal probability distribution function

w = Wetherill designation for a current estimate of the median

$\bar{w}$ = average of the w's

X = Einbinder's notation for a "positive" response

x = often designates a random variable, but is used in sensitivity analyses to denote the stress or stimulus level

$xi = x_i$ = stress level

x_c = designation of a stimulus level by Ross (see Eq. 9-70)

x_s = transformed stress

x_1, x_2 = different levels of the stress

x_α = percentage point giving probability level α

$Y = F^{-1}( )$ = designates inverse transformation of function F

$y = z + 5$ = probit p = a probit for the normal model

$y_i = x_i - \gamma$ = transformation for the Weibull model

y_s = transformed responses

y_0, y_1, y_2 = different values of the argument y for $p = F(y)$, i.e., the transformation

$z_i = z(x_i)$ = standardized normal deviate

0 = Einbinder's notation for a "negative" response

0 = "negative" response = an initiation in Table 9-4

1 = "positive" response = no initiation in Table 9-4

$\alpha = \mu/\sigma$ for the logistic model

α = parameter of the logistic model

α_0, β_0 = initial estimates for an iteration
α_1, β_1 = first iterated estimates, etc.
β = parameter of the logistic model
β = Weibull model shape parameter
$\beta = 1/\sigma > 0$
γ = Weibull location parameter or start of frequency
γ_R = starting frequency point for a reflected Weibull model
$\Delta\alpha$ = small change in α
$\Delta\beta$ = small change in β
$\Delta\sigma$ = small change in σ
δ_i = random variable which takes on the value zero or one
$\theta = \sigma^{1/\beta}$ = parameter of Einbinder for the Weibull model
μ = population mean, usually of a normal population
σ = population standard deviation, usually a normal universe
σ = Weibull model scale parameter
σ_y = denotes standard deviation of the subscript y whatever y represents, $\sigma(\ )$ also used
$\sigma_{\hat{\sigma}}$ = standard error of the $\hat{\sigma}$
$\hat{}$ = denotes estimate of

(There is some special notation used by DiDonato and Jarnagin or Einbinder in Computer Programs 9-1 and 9-2, which is not listed here. However, wherever possible, we have endeavored to use the authors' notations for the key parameters, as described in the text.)

9-1 INTRODUCTION

As contrasted to the topics discussed so far in this handbook, the statistical analyses of sensitivity-type experiments represent some very different methodologies in which the Army analyst may desire expertise. Nevertheless, sensitivity experimentation and the associated special statistical analyses are quite important in their own right. Such procedures are required, for example, in penetration of armor studies, the analysis of the sensitivity of primers or explosives, dosage-response curves, bioassay experimentation and analyses, dosage-mortality curves, quantal response curves, radiation-mortality curves with risk analyses of people, and time-response or time-mortality curves. Thus our label "sensitivity analysis" is merely a fairly well-accepted Army term that has come into some prominence due perhaps to explosive sensitivity or to the penetration of armor studies to determine the ballistic limit of armor plate and the apparent desire to distinguish it from the long-existing field of bioassay.

During World War II, our country had a major problem relating to tests for the acceptance of armor plate to be placed on tanks for personnel protection. An important analytical task in this connection was to determine the penetration limit of armor plate fired at with armor-piercing (AP) rounds for the purpose of estimating the ballistic limit of the plate. The ballistic limit, or the V_{50}* as it came to be known, developed along with it the definition that V_{50} would be the striking velocity for which 50% of the AP projectiles would penetrate the plate. As is well-known, all projectiles fired from weapons exhibit random variation in velocity caused by slight variations in the amount (weight) of propellant loaded into the cartridge case, the random position of the propellant in the case when firing occurs, some variation in ignition properties, etc. For even a constant level of striking velocity against the plate, it is found that only a fraction of the projectiles might penetrate, depending on the velocity level. At some "low" velocity level, no projectiles will penetrate the armor plate; while at some "high" level of velocity, one might expect that 100% of the projectiles will penetrate. However, there will be cases for which some percentage of the striking projectiles will break up and not penetrate, even for the higher velocities, especially for "sloped" armor or armor plate at the higher angles of obliquity. Thus, and somewhat in summary, it is reasonable to expect that a lower velocity will exist for which

*The correct label would be $V(50\%)$ or $V_{0.50}$.

there will be 0% penetrations and the percentage of penetrations will increase up to 100% for some minimum higher velocity. The zone in between has often been called a "zone of mixed results", but somewhere in the middle is the V_{50} or ballistic limit. Clearly, it becomes quite difficult to determine the lower and upper endpoints of the zone of mixed results since either a penetration or a nonpenetration occurs in firing, i.e., a "quantal" response, and a large number of rounds must be fired to estimate the 0% or the 100% penetration levels with any precision or accuracy. Indeed, with quantal responses, i.e., "all or nothing" responses, even the estimation of the median striking velocity V_{50} for 50% penetrations would be difficult enough with small sample sizes. An added problem is that the zone of mixed results might extend over several hundred feet per second, or even a thousand feet per second, and the standard deviation in velocity level of armor projectiles fired could easily be 10, 15, or 20 ft/s. Thus the striking velocity against the armor plate cannot be controlled very precisely either. For an assumed cumulative normal distribution of the proportions of penetrations over the zone of mixed results, therefore, the standard deviation of the curve can be expected to be as large as a hundred feet per second or perhaps several hundred feet per second.

Similar considerations apply to other Army sensitivity analyses, including the sensitivity of explosive to shock, or the comparative sensitivity of primers, etc., although the height of drop onto such devices can be controlled rather accurately. In any event, whatever analytical methods we develop will apply equally well to bioassay-type problems, dosage-mortality curves, or quantal response studies, whatever the field of application. Our major point concerns the urgent need for small sample sizes, especially since most Army tests are destructive in nature. Before proceeding, we must say that the assumption of only a normal distribution for the cumulative percentages of penetrations is not always tenable, so that we must often consider the possibility of applying other models, including the use of nonsymmetric distributions, such as the Weibull or logistic laws or models. We might add that we will primarily be interested in the estimation of the location and scale parameters of the distribution of sensitivity results even though the endpoints of 0% occurrences and 100% occurrences are quite critical in many applications.

To acquire the proper understanding of the more basic problems in sensitivity analysis, we will formulate the approach in terms of some analytical procedures, which help depict what is really taking place.

9-2 BRIEF ANALYTICAL FORMULATION OF SENSITIVITY ANALYSES

For the treatment of later estimation problems, our discussion starts with a rather general probability density function (pdf), which we will call $f(x)$. The pdf may take on any of several different forms of interest in Army applications. For example, often there will be the need to analyze sensitivity-type data, which follow the normal density, or

$$f(x) = (1/\sqrt{2\pi})\exp[-(x-\mu)^2/(2\sigma^2)]^* \tag{9-1}$$

where the population mean μ is also the median or 50% striking velocity, dosage, etc., and the scale parameter σ gives the measure of the width of the zone of mixed results since it is the standard deviation. Thus for all practical purposes the expected width of this zone would be about 6σ for the assumption of a normal distribution. We must hasten to point out the exact nature of a quantal response. For example, suppose that we are firing AP projectiles at armor plate, and the striking velocity is represented by the variable x, which, for illustrative purposes, we set equal to 1000 m/s. Even though our problem is to estimate the V_{50} or μ and the standard deviation σ, let us assume that $\mu = 800$ m/s and that $\sigma = 100$ m/s.** Then, by designating the unit or standard normal variable z, we see that

$$z = (x-\mu)/\sigma = (1000 - 800)/100 = 2. \tag{9-2}$$

This means that under the specified firing conditions the chance that a response, in this case a penetration, occurs is

$$p = p(x) = F(x) = \int_{-\infty}^{2} f(z)dz = 0.977 \tag{9-3}$$

*We use x for a general variable to represent the striking velocity, height of drop onto an explosive, a dosage level, stimulus, etc.

**For a normal population, the mean $\mu = x_{0.50}$.

Therefore, the actual firing of the AP round with a striking velocity of 1000 m/s, either a penetration or a nonpenetration would occur, but the chance for a penetration of the armor is very high indeed, i.e., 98%. Put another way, if a very large number of AP projectiles with a striking velocity of 1000 m/s were fired against the same plate, approximately 98% would penetrate the plate and 2% would not. As we have indicated, however, our problem is to take the penetration and the nonpenetration results with their particular striking velocities and to estimate μ and σ.The reader will immediately recognize that we need to develop a method of test that will more or less guarantee the minimum number of striking velocities, which will render a mixture of penetrations and nonpenetrations from which the μ and σ may be estimated with precision. This is called a strategy. In fact, the problem of developing the "best" strategy in some sense turns out to be rather critical in sensitivity analyses, such as the determination of the ballistic limit. If we are concerned primarily with the estimation of the mean μ of the normal distribution assumed, it would appear wise to shoot with those striking velocities that give about equal numbers of penetrations and nonpenetrations. If we also want to estimate the standard deviation of the distribution, then it would appear wise to go somewhat away from the center of the distribution because the standard deviation reaches out to the point of inflection of the normal curve. Finally, if our interest were primarily to estimate a level of some very small percentages of penetrations, our strategy should involve converging on such a small fraction. A similar problem applies to the estimation of a very high level of successful penetrations of the armor. For the mean and standard deviation, and estimation thereof, the "up and down" strategy of Dixon and Mood (Ref. 1) seems appropriate and has gained wide acceptance. For the up and down procedure, and for tests of armor, the striking velocity is increased if a nonpenetration occurs, and the striking velocity of the next round fired is decreased if a penetration occurs—thus the term up and down. This strategy keeps testing near the middle of the distribution although the problem of starting the test at a good level remains, and one needs to know the best interval at which to change the striking velocity. For a normal distribution the best interval d is such that $2\sigma/3 < d < 3\sigma/2$ (Brownlee, Hodges, and Rosenblatt, Ref. 2). With this spacing, Brownlee, Hodges, and Rosenblatt (Ref. 2) found that small samples will give an efficient estimate of the median dosage, or here the V_{50} striking velocity, i.e., $x_{0.50}$.

We will delve more into the use of various strategies in the sequel, but for the present our main purpose is to continue with two other useful models for Army applications, namely, the Weibull and the logistic models as contrasted to the normal.

For the Weibull and logistic models, simplicity is attained by expressing their analytical form as cumulative distributions so that the corresponding pdf's may be obtained by differentiation. Hence the cumulative distribution of the Weibull model for sensitivity analyses would be taken as

$$F(x) = 1 - \exp[-(x-\gamma)^{\beta}/\sigma] = 1 - \exp\{[-(x-\gamma)/\sigma^{1/\beta}]^{\beta}\} \tag{9-4}$$

where

γ = start of the frequency
β = shape parameter
σ = scale parameter.

If we deal with the two-parameter Weibull model instead of the three-parameter one in Eq. 9-4, the start of the frequency is at zero, and hence $\gamma = 0$. The reader is aware that the Weibull model can take on a variety of shapes and is more general than the symmetric normal distribution. A number of authors have in recent years become very interested in the Weibull model. See Einbinder (Ref. 3) and others in the references and bibliography for the use of the Weibull model in sensitivity analyses.

Undoubtedly there are a rather large number of Army applications for which the logistic model is applicable, especially perhaps in bioassay-type problems. The logistic model is represented by

$$F(x) = \{1 + \exp[-(\alpha + \beta x)]\}^{-1} \tag{9-5}$$

where in terms of the parameters α and β, the mean of the logistic distribution is

$$E(x) = \mu = -\alpha/\beta. \tag{9-6}$$

The reader will thus understand that the logistic curve—like the normal curve—will range over the limits from minus infinity to plus infinity, but on the other hand it will take on a variety of shapes between its limits. To date, it does not seem that the logistic distribution has been used very widely in Army engineering-type applications.*

The normal, Weibull, and logistic models represent the basic three types of distributions we will cover in this chapter although others, such as the gamma or exponential distributions, also could have extensive applications. (The Weibull model includes the exponential model as a special case.) Even though we have indicated the three models we will discuss, there is yet another very important consideration to be brought forward concerning sensitivity analyses before we proceed any further, i.e., the rather indirect method of estimation of the parameters that is required.

If we observe the normal model of Eq. 9-1, the Weibull model of Eq. 9-4, and the logistic model of Eq. 9-5, note that $F(x)$ is the cumulative distribution function and hence gives the chance of a response at the level of stimulus x. Hence if we designate that the levels of response are $x_1, x_2, \ldots, x_i$, the probability of a response at level x_i is notationally

$$p_i = p(x_i) = F(x_i) \tag{9-7}$$

where we use p simply to mean probability. Now in a test of a "specimen"—whether it be armor plate, an explosive, a primer, etc.—we will observe either a "response" (penetration) or "no response". That is, the observed random variable is either a one or a zero; "one" represents a response and "zero" the lack of any response. Thus we may look upon the sensitivity experiment as did Golub and Grubbs (Ref. 4), who pointed out that a random variable δ_i could be considered that takes on a zero value or a one for each level of stimulus, so that the likelihood of occurrence of the observed sample, or the chance of the observed set of observations, is given by

$$P = \prod_i (p_i)^{\delta_i}(1 - p_i)^{1-\delta_i} \tag{9-8}$$

where we take the product Π with respect to a series of observations, $i = 1, 2, 3$, etc., to range over any number of trials we may want to include in the experiment for our particular estimation problem. Hence we have used the concept of δ_i simply to denote the actual observational responses. When $\delta_i = 1$, a response has actually occurred even though its probability of occurrence is p_i (which may take on any value between zero and one), and when $\delta_i = 0$, there is no response with probability of occurrence equal to $q_i = 1 - p_i$. The mean value of δ_i is $E(\delta_i) = p_i$, and the probability that $\delta_i = 1$ is $Pr(\delta_i = 1) = p_i$.

Recall at this point that p or p_i is a cumulative distribution—whether it be the normal integral of Eq. 9-1 up to a value x_i, the cumulative Weibull given in Eq. 9-4, or the logistic form in Eq. 9-5—so that the estimation of the indicated parameters may become somewhat cumbersome to say the least. In fact, it can be seen that one approach would be to take logarithms of the likelihood indicated by Eq. 9-8 and to proceed with Fisher's principle of maximum likelihood (ML) estimation. In fact, this is often just what is done. Some readers may wonder why we have formulated the sensitivity analysis problem in terms of the rather general but particular response model as given in Eq. 9-8. The answer is that a solution for estimation of the unknown parameters based on Eq. 9-8 would apply to a wide variety of practical problems for which small sample sizes are more or less mandated. Moreover, in the case of firing at armor plate, one cannot launch a projectile at any desired velocity level. Rather he may aim for 1000 m/s, but due to random variation in muzzle velocity, the striking velocity at the plate may be, for example, 990 or 1008 m/s, and p_i is general.

With this discussion, we have reached the stage at which it seems advisable to discuss some test or firing strategies often followed in sensitivity analyses. After that, we will go briefly into the problem of estimation of parameters.

9-3 SOME USEFUL TEST STRATEGIES

Perhaps the more useful test strategies for many Army applications include the complete rundown test, the up and down test of Dixon and Mood (Ref. 1) developed in connection with explosive sensitivity-type

*Use of the logistic distribution is about equivalent to using the normal model. (See par. 9-3.) The normal model is often called the "probit" form, and the logistic, the "logit" form.

investigations, the Langlie (Ref. 5) one-shot test, the Robbins-Monro stochastic approximation method (Ref. 6), the one-shot transformed response test, and other transformed response strategies. All but the first of these testing strategies may be referred to as sequential sensitivity tests, which do not involve fixed or preset sample sizes for the tests, and it becomes desirable to use some kind of stopping rules along with them whenever a suitable number of tests have been attained.

9-3.1 THE COMPLETE RUNDOWN TEST

In the complete rundown test, the idea is to test a fixed number of items at each level over the estimated zone of mixed results so that the percentage of responses will vary from near zero to 100%. For example, this has often been a natural test in primer sensitivity studies. In this application, perhaps 50 primers are tested at each inch of drop height from the level where nearly all 50 explode down to a low height for which none of the 50 function. With the percentage of responses so varying, one may fit a curve, for example, by the method of least squares to the observed fractions of responses to summarize results.

9-3.2 THE UP AND DOWN TEST OF DIXON AND MOOD*

For the up and down test strategy, designed by Dixon and Mood (Ref. 1) to estimate the mean and standard deviation of the normal distribution, the basic idea is to increase the level of stimulus when the test specimen does not respond and to decrease the level of stimulus by a step when the test specimen does respond. The Dixon and Mood up and down test strategy is indicated on Fig. 9-1, where an "X" means a response and a "0" means no response. The initial test level is at the stimulus level that represents the best estimate of the 50% point of the distribution. The true value of the 50% point is hardly ever known; we are, in fact, testing to establish it; therefore, one has to make a wild guess at first to get started. The step size also is fixed and must be set in advance and the best value is about one standard deviation, as indicated in par. 9-2. Clearly, the up and down test procedure concentrates the observations very near the mean—just where they should be for the normal distribution. However, the up and down strategy does not do very well for the problem of estimating the extreme percentage points of a distribution—unless the curve is in fact a normal one, and the mean and standard deviation are determined quite accurately. It has been claimed that the up and down test may be too sensitive to the starting level of stimulus and the step size although the nature of sensitivity analyses is such that in many applications little is known about the true location of the underlying distribution and its shape! Moreover, if one has to rule out a very large number of tests, but still is interested in the general nature of the phenomena studied, he may want to use the up and down strategy, at least initially.

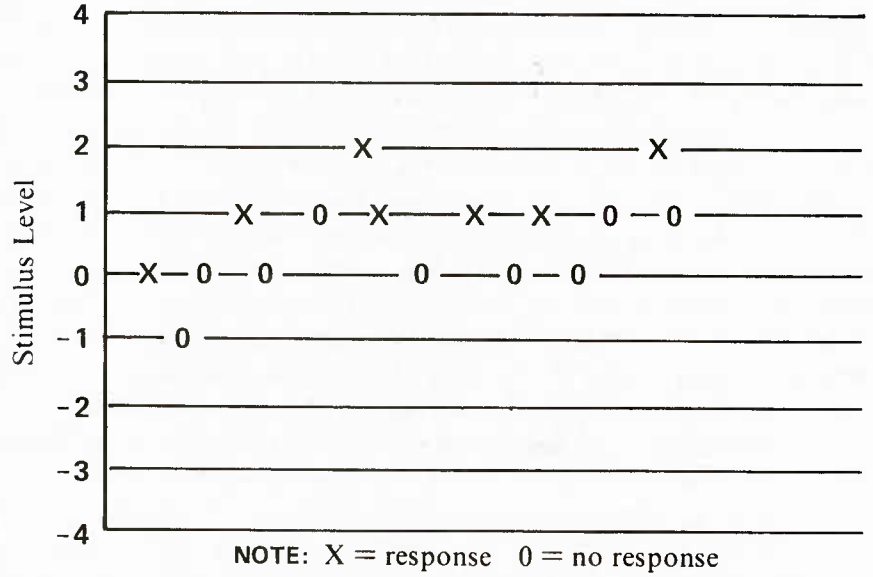


Figure 9-1. A Typical Up and Down Experiment

*The up and down method of testing has often been referred to as the Bruceton method.

9-3.3 THE LANGLIE ONE-SHOT STRATEGY

Langlie (Ref. 5) suggested a sequential test strategy that was to overcome certain of the difficulties associated with the up and down procedure. In fact, the Langlie test strategy also makes use of continuously variable stress levels and was suggested to be insensitive to the starting level and the *a priori* choice of the step size. It does, however, depend on estimates of the endpoints of the zone of mixed results that apparently are obtained most often from engineering considerations. Also these endpoints may come into play during the selection of the next level of stress in the test. Some analyses have indicated that the Langlie test strategy may be more efficient than the up and down procedure for estimation of the location and scale parameters of the normal distribution although some further comparisons should be made. We note in passing that Langlie labeled his strategy as involving a reliability test method of one-shot items. In this connection, Langlie apparently visualized items operating in some region of the environment for which the stress levels were such that all items were supposed to operate satisfactorily. However, for the higher and increasing stress levels, the items would begin to fail. In fact, there would be a distribution of failures on the stress scale. As Langlie stated in Ref. 5, "In the case of specimens having extremely short lives, it is possible only to anticipate a stress level and then operate the specimen under this environment to see whether or not it is successful. Such items are referred to as 'one-shot' items. Examples of 'one-shot' items include short duration rocket motors, switches, relays, and a host of similar items. Each part, when tested, will function satisfactorily or unsatisfactorily; such an 'all or nothing' situation is referred to . . . as a 'one-shot' test.". Hence it is seen that Langlie's one-shot label simply refers to the test of an item in the ordinary sensitivity test procedure and not to something otherwise quite special.

Perhaps the best way to illustrate Langlie's test strategy is to give his own example, which presents some results on the test of thermal batteries, as depicted in Fig. 9-2. The purpose of Langlie's actual one-shot test on thermal batteries was "to determine the reliability with regard to high temperature". In this instance, the batteries were designed to perform reliably at 145° F. Langlie indicates that on the basis of conservative engineering judgment and some limited development test data: (1) the lower limit on temperature was selected to be 100° F—the level at which all thermal batteries would be expected to perform satisfactorily—and (2) the higher temperature limit was selected to be 350° F—the level at which all thermal batteries would be expected to fail. Thus stress was taken to be the temperature level, and "Once the test level and failure criteria have been established, the test commences by selecting the first stress level at the midpoint of the interval.". Therefore, the first battery is tested at the temperature of $(100 + 350)/2 = 225°$ F, and Langlie records a "1" (in the right-hand column of Fig. 9-2) if there is a "positive" response, which in this case means a failure of the battery, and a "0" if the response is "negative", i.e., the battery operates satisfactorily.

The general rule of Langlie for obtaining the $(n + 1)$st stress level, i.e., after n trials, is to work backward in the test sequence, starting at the nth trial, until a previous trial (call it the pth trial) is found such that there are as many successes as failures in the pth through nth trials. The $(n + 1)$st stress level is then obtained by averaging the nth stress level with the pth stress level. If, however, there exists no previous stress level satisfying the requirement just stated, the $(n + 1)$st level of stress is obtained by averaging the nth stress level with the lower or the upper stress boundary of the test according to whether the nth test result was a failure or a success, respectively. To illustrate the second stress level, it is noted that the first test at 225° F resulted in a battery failure, and it is not possible to find any previous stress level in the test where all intervening results even out. Therefore, for the second stress level we take it equal to $(100 + 225)/2 = 163°$ F since we clearly must go to a lower temperature to search for a successful battery operation. The result is a zero or success.

For the third stress level we have a zero in the second test and a one for the first trial, so the numbers of failures and successes are equal. Hence we simply average 225 and 163 to obtain the temperature of 194° F for the third trial, and the result of the next one-shot test is a successful battery operation, i.e., a zero.

The process continues. For example, observe the eighth shot or stress level. We note in this particular case that the immediately preceding tests, or the 4th through the 7th (but not the last two or three tests), give two positive and two negative responses—an equal number—and hence for the 8th stress level we average the 4th and the 7th and obtain $(178 + 227)/2 = 203°$ F.

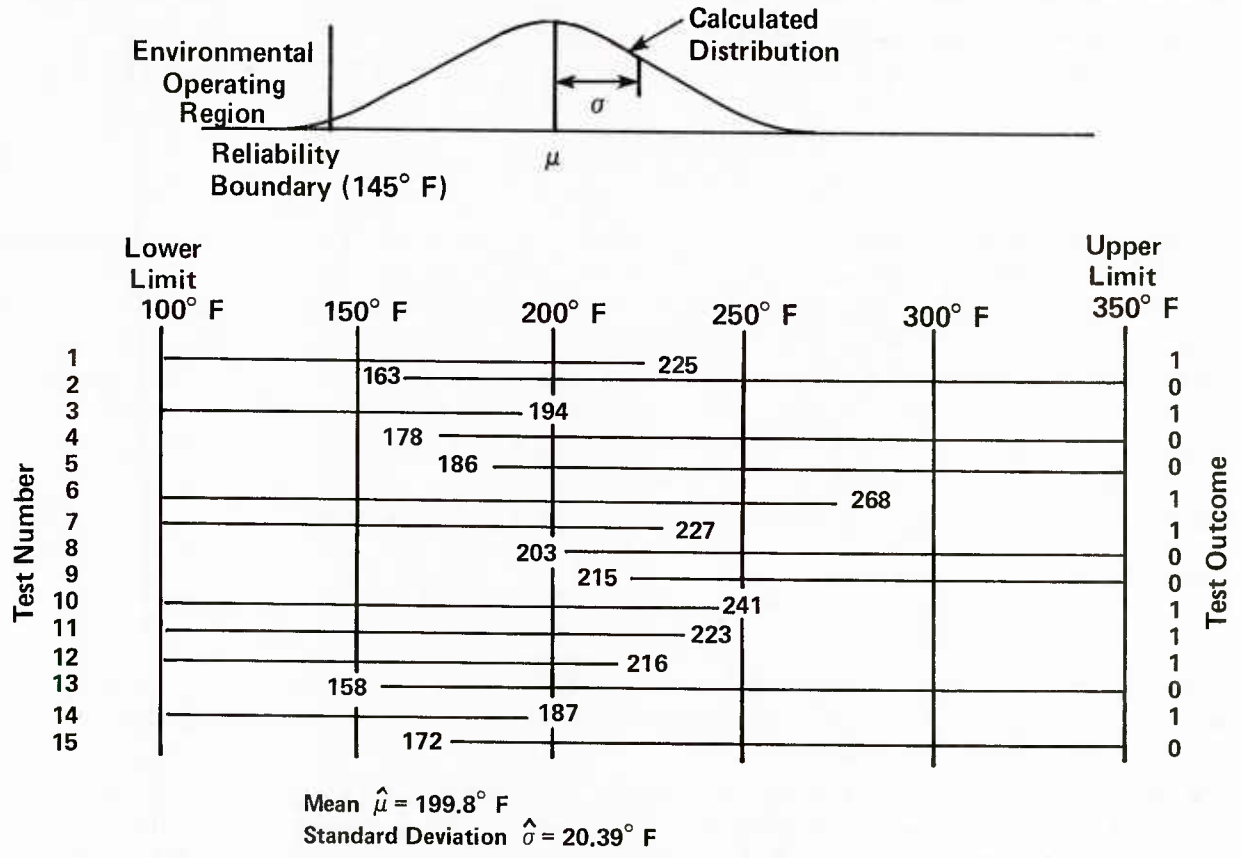


Figure 9-2. "One-Shot" Test to Determine Failure of Thermal Batteries (Ref. 5)

Finally, if no previous level exists to satisfy this criterion, the very last stress is averaged with either the upper or lower limit (this depends upon whether an increase or a decrease in stress is required). This should be sufficient to give the reader a good idea of the proposed Langlie test strategy, and this procedure more or less concentrates the test results near the median although occasionally the last stress level may have to be averaged with one of the original limits.

The actual process of estimating the mean and standard deviation of the assumed normal distribution by Langlie involves Fisher's ML estimation, which we discuss in par. 9-4. Note on Fig. 9-2 that the ML estimates of the mean and standard deviation are, respectively, 199.8°F and 20.4°F.

Thus as we have said, both the up and down and the Langlie strategies tend to concentrate the test results near the central part of the sensitivity distribution, and they are efficient for estimating the mean and the standard deviation of, for example, the assumed normal distribution of responses. However, there is another efficient procedure for estimating the median of the response curve, i.e., the Robbins-Monro stochastic approximation process (Ref. 6), which has been studied rather thoroughly in a key paper by Wetherill (Ref. 7). In fact, Wetherill's paper covers a rather extensive study of a number of possible strategies for estimating not only the median or $x_{0.50}$ of the response distribution, but also percentage points such as a lower one, e.g., $x_{0.05}$ or an upper one, $x_{0.95}$. Note that if we are dealing with a normal distribution, the estimation of $x_{0.84} - x_{0.50}$ would give the standard deviation.

9-3.4 THE ROBBINS-MONRO STOCHASTIC APPROXIMATION METHOD

In 1951 Robbins and Monro (Ref. 6) introduced a very general method of stochastic approximation for the regression-type situations, and hence it applies to the sensitivity analysis problem. Suppose in the quantal response type of endeavor we want to estimate any general percentage point or probability level p in terms of the value of the response level x that results in such a desired probability. Then the Robbins-Monro procedure

means that in a series of observations, such as $\delta_i = 0$ or 1 taken at levels of stimulus x_i, one uses for the next trial a stochastic approximation to obtain x_{i+1} based on

$$\hat{p} = x_{i+1} = x_i - a_i(\delta_i - p) \tag{9-9}$$

where p is the probability level desired, for example, 0.50, and the quantity a_i is a series of constants chosen to depend on i in a manner that successive changes in the level become smaller and the observations converge to the true value p desired. It would appear, based on a theorem of Chung (see Ref. 8), that the best asymptotic function for a_i is a constant divided by i, i.e., c/i, c being the constant. This function is best in the sense that it produces the most rapid convergence of the estimator to the value of p desired when the x_i are linearly related to the p_i. Hence it can be said that the series of constants c/i will be best for the quantal response problem in an asymptotic sense since locally linearity will be the case. In Ref. 7 Wetherill refers to the Robbins-Monro process as "Routine 1" and the up and down strategy of Dixon and Mood as "Routine 2".

Wetherill (Ref. 7) points out that the asymptotic variance of x_{n+1} is given by

$$\mathrm{Var}(x_{n+1}) = 1/[\beta^2 p(1-p)n] \tag{9-10}$$

for the logistic model of Eq. 9-5. Therefore, the variance of the estimated stimulus to obtain a probability level of p depends on p in the denominator, its complement from unity, the size of the slope in Eq. 9-5, and the number of iterations made in the process.

As a result of some theoretical investigations and many Monte Carlo simulations, Wetherill (Ref. 7) indicates that the Robbins-Monro stochastic approximation procedure is very efficient for the estimation of the stimulus level for the median, or $p = 0.50$, both as a method of placing observations properly and as a technique of estimation itself. The Robbins-Monro stochastic approximation procedure is very "robust" to errors in starting values of the sequence and also to the value of the constant c. (See Ref. 7 for a discussion of the optimum values of the constant c.) Moreover, actual small sample variances closely follow the asymptotic values such as those given in Eq. 9-10. However, Wetherill does conclude that the Robbins-Monro strategy is really unsuitable for estimating even moderately extreme values of the stimulus level that result in $p = 0.25$, for example. This would mean that estimation of the stimulus level for $p = 0.05$, 0.01, or 0.99 would be expected to give all kinds of trouble, so we see the great difficulty in the sensitivity problem. In fact, it would appear that unless one is willing to conduct an enormous number of trials, he must often be content with imprecise estimates of the location and scale parameters.

Finally, for the Robbins-Monro technique Wetherill indicates in Ref. 7 that the procedure is "asymptotically fully efficient for estimation of any p, in the sense that it has an asymptotic variance equal to the minimum attainable variance . . . However, this conclusion only holds if the optimum value of [the constant] c is used, which depends on the slope of the response curve at p. Since c must be chosen in advance, and Routine 1 provides almost no information about slope, then loss of efficiency will result.". Thus, although the Robbins-Monro procedure may not be very sensitive to the value of c for estimation of the stimulus giving $p = 0.5$, there will be percentage points for which the efficiency of the technique may depend markedly on the constant c.

Although we are concerned primarily with the Robbins-Monro stochastic technique in this paragraph, Wetherill goes into a rather extensive evaluation of the Dixon-Mood up and down procedure for $p = 0.5$ since it would be a natural competitor of the Robbins-Monro method. In this connection, Wetherill (Ref. 7) states for the up and down method that "If a spacing of between 1.5σ and 2.5σ units is used, the asymptotic efficiency of Routine 2 is about $20/27 = 74\%$. This efficiency drops off sharply for values of the ratio of spacing interval to slope constant β [for the logistic model] outside this range. In most practical situations the slope is not known, so that the asymptotic efficiency of Routine 2 depends critically on one of the unknown parameters.". In fact, the asymptotic efficiency of the up and down method depends rather more heavily on the choice of the spacing interval than the Robbins-Monro technique depends on the choice of the constant c. On an overall

basis Wetherill concludes that the up and down strategy may be very highly efficient (about 80% or better) for estimating the median stimulus or dosage, provided there is a good choice of the spacing steps, but that considerably improved efficiency could be effected with a division of the spacing at suitable points. Moreover, one must assure a good starting level for the up and down strategy; if not, there would be some adverse effects.

The reader will very likely note that this discussion of efficiency is restricted to the up and down and the Robbins-Monro procedure comparisons and that a study of the relative efficiency of the Langlie strategy with these two has not been given. Thus it can be said that much needed research remains to be done, and a computer likely would be required to conduct many of the necessary comparisons through Monte Carlo-type experiments. Also for anyone interested in conducting further research on the sensitivity analysis problem, Wetherill's paper should be considered mandatory reading since it includes stopping rules.

Finally, as a remark or two concerning the estimation of low or high levels of p, Wetherill (Ref. 7) expressed some surprise that sequential strategies to estimate p when $p \neq 0.5$ have not attracted more attention. Perhaps, however, this lack of emphasis merely indicates the difficulty of the problem! In all, Wetherill investigated some 15 strategies for all levels of p studied but found that for estimation of the stimulus level for $p = 0.95$, some of the most favored strategies gave large expected squared errors, large biases, and very frequent samples producing extrapolated estimates. He even recommended that the estimation of extreme percentage points should be "avoided at present". Nevertheless, Wetherill's investigations have had a very decided impact on the transformed response strategies, which were advanced by Einbinder (Ref. 3) and which, in fact, combine the Langlie strategy with the better of the Wetherill routines or strategies studied as we will see next. Therefore, we will go into a discussion of some proposed strategies for the high and low percentage points.

9-3.5 THE ONE-SHOT TRANSFORMED RESPONSE TEST STRATEGY (OSTR)

The percentage points in the tails of distributions have very important practical implications and are required in the design of products. For example, in the design of armor for a tank, the design engineer may have a good idea concerning the highest striking velocity of enemy antitank projectiles. Thus if he knew the velocity level at the 0.1% point for the lower end of the percent penetrations vs striking velocity curve, the armor thickness could be determined so that practically no enemy projectiles would defeat the tank. If we are interested in estimating stimulus levels for the *lower* tail area of a distribution, it would appear that the test strategy should be such that rather rapid convergence to the required probability level is assured. This would mean that the stress levels should be taken in a manner that would make it easier to *decrease* the stress than to increase it. Of course, for the estimation of the high percentage points, the reverse should be effected. Einbinder (Ref. 3), based on the work of Wetherill (Ref. 7), has suggested a response transformation to bring about such results. In the course of his key study, Wetherill (Ref. 7) noted that occasionally some very peculiar sequences of outcomes would occur, such as a series of ones or zeros, which, when continued, would provide very little or no information about the response distribution. Accordingly, Wetherill suggested using a change of response stopping rule rather than a fixed sample size to minimize the loss of information. In fact, among the many strategies or routines studied by Wetherill, one in particular took cognizance of this (Wetherill's Routine 15) and was based on a form of inverse sampling. We quote from Wetherill's paper (Ref. 7):

"Routine 15 (Inverse Sampling): Use a fixed series of equally spaced levels and after each trial, estimate the proportion p' of positive responses at the level used for the current trial and consecutive with it, that is, back to the last change of level. If $p' > p$ and p' is estimated on n_0 trials or more, decrease the level one step. If $p' < p$, increase the level one step. If $p' = p$, make no change in level. . . . Routine 15 is to work as follows: for $p = 0.75$ (for example) estimation choose $n_0 = 4$ and make no change of level for positive responses until four consecutive positive responses have occurred, then move the level one step down; increase the level one step if successive results after a change in level are 0, or 1-0, or 1-1-0; for 1-1-1-0 make no change in level but move according to the next response.".

Hence for Wetherill's Routine 15, we see that when we are trying to estimate an upper percentage point, the level may be increased as soon as a nonresponse occurs in four consecutive trials thereby forcing the testing up near the desired percentage point $p = 0.75$. Wetherill (Ref. 7) thus proposed a stopping rule based upon a specified number of changes of the response type instead of a fixed number of trials total.

Einbinder (Ref. 3) suggests a strategy that combines such a feature of Wetherill with the Langlie test strategy and calls the quantile around which the test levels tend to concentrate for a given n_0 the "transformed median percentage (TMP)". The overall test strategy is labeled the "One-Shot Transformed Response (OSTR)" procedure by Einbinder (Ref. 3). Perhaps the OSTR is illustrated best by an example of Einbinder given in Table 9-1. He considers a case for which the lower and upper Langlie-type boundaries are taken to be zero and 70, respectively, and a sequence of trials to estimate the stimulus level for a probability equal to $(0.50)^{1/3} = 0.7937$, which is the transformed median response (TMR) based on a Wetherill $n_0 = 3$. Note in this connection that the $F(x)$ based on the assumption of a normal distribution in Eq. 9-1, or the Weibull in Eq. 9-4, or the logistic in Eq. 9-5 each gives the chance of a positive response. Thus for $n_0 = 3$ the chance of a downward change in direction under Wetherill's inverse sampling scheme is

$$P = Pr = [F(x)]^{n_0} = (0.7937)^3 = 0.5 \tag{9-11}$$

and such a probability becomes greater for the higher percentage points, such as 0.90 or 0.95 unless the n_0 is changed. Thus the procedure may be adjusted to conform to almost any high percentage point or to the lower percentage points as well. In effect, therefore, due to the particular sequences required before a change in the level of stimulus, one can—by proper choice of the n_0 along with the percentage point p desired—conduct a sensitivity test and analysis so that he is more or less "aiming for a median value".

We now proceed to a discussion of the Einbinder OSTR test strategy (refer to Table 9-1). Recall that we want to carry out a strategy to estimate the stimulus or stress x that gives the 79.37% point of the cumulative distribution. The first "shot" is then taken at the stress level of $0.7937(70) = 56$ for an upper Langlie boundary of 70, and the response is a "success", i.e., "1". (Einbinder, Ref. 3, uses either a "1" or an "X" for a positive response.) Now since $n_0 = 3$, under the Wetherill rule we continue with the same stress $x = 56$ for the next "shot", which is also a positive response, i.e., 1. Again we take the next test at $x = 56$ since we got a 1, and the result is a third positive response. This third positive response indicates that we must go down in stress level, however. Therefore, we must now take the lower Langlie boundary of zero and average it with the 56 to obtain the next stress level of $x = 28$ for which a positive response is still obtained. Since we have only a single positive response at the stimulus level of 28, we should take the next "shot" at that same level, and we obtain a nonresponse, i.e., 0. This means, therefore, that we must increase the stress level, and we also note and record that this "up" brings about the first change number. Moreover, the average of the two levels of 28 and 56 gives Wetherill's first $w = 42$ as the first estimate of the 79.37% point. The next stress level, i.e., for the 6th shot, is 42.0, and a nonresponse is observed, which means the stress level must be increased. At this stage we have two U's and one D, or unbalanced responses, so that by using the upper boundary of 70 with the last level 42, trials are continued at $x = 56.0$. The experiment continues as indicated on Table 9-1, and the fourth change of response occurs at the 16th trial. A change of response type is said to occur whenever an alternation of the response is obtained. Wetherill's stopping rule (Ref. 7) is based upon a specified number of changes in response type rather than any fixed total for the number of trials. The number of observations or trials in an experiment of this type results in a random variable for Wetherill's stopping rule, and the expected sample size for a particular number of changes or responses will increase with the parameter n_0 or the farther out in the tails of the distribution we desire testing to take place. Moreover, for each sequence of trials on the transformed scale that represents a change of response, a reasonable estimate of the 50% point or 50th percentile is the midpoint of the stress interval in which the change took place. We denote these estimates by w, due to Wetherill who proposed such a rule for the Dixon-Mood up and down method to close in on the fineness of the interval instead of sticking to equal spacing. Also the application of the Wetherill inverse sampling strategy to the Langlie technique clearly would seem to be very efficient and accurate. As pointed out by Einbinder (Ref. 3), each change of response for the proposed strategy results in a separate estimate of the transformed 50% point, and the overall average $\bar{w}$ of Wetherill is taken as the expected transformed median.

For the example of Table 9-1, the average value $\bar{w} = 50.09$ after the fourth change number or 16th trial is the estimate of the 79.37 percentile of the underlying distribution. The reader should observe that Wetherill's $\bar{w}$ is a simple estimate to calculate, especially compared to ML estimates. Moreover, such simple estimates could well be used for starting values in the ML estimation of par. 9-4, for example.

TABLE 9-1

OSTR TEST FOR $n_o = 3$, TMP = 0.7937 (Ref. 3)

Lower Langlie Boundary, $A = 0$. Upper Langlie Boundary, $B = 70$.
For the first trial, $i = 1$, take the stress = 0.7937(70) = 56.

Trial i	Stress x_i	Response $\delta = 0,1$	Response Type D or U	Change Number	Wetherill's w
1	56.0	1			
2	56.0	1			
3	56.0	1	D		
4	28.0	1			
5	28.0	0	U	1	42.0
6	42.0	0	U		
7	56.0	1			
8	56.0	1			
9	56.0	1	D	2	49.0
10	49.0	1			
11	49.0	0	U	3	52.5
12	52.5	1			
13	52.5	0	U		
14	61.25	1			
15	61.25	1			
16	61.25	1	D	4	56.875
					200.375

D = Down: 111
U = Up: 0, 10, 110

$\bar{w} = 200.375/4$
$= 50.09$

9-3.6 TRANSFORMED RESPONSE STRATEGIES FOR GENERAL n_o

In connection with transformed response strategies for any value of n_o, Einbinder (Ref. 3) has developed a table of characteristics of some of these typical strategies. That table is included as Table 9-2. The upper and lower tail areas that are estimated and around which the Wetherill-Langlie strategy or one-shot test levels tend to concentrate also are given in the last two columns of Table 9-2. The table may be extended to any n_o and/or transformation desired by the experimenter. In Table 9-2 we use Einbinder's X to denote a positive response and a 0 to denote a negative response; these apply as noted for probabilities or percentiles $p > 0.5$. For the lower tail areas of distributions of interest, we must redefine the responses so that 0 represents a positive response and 1 or X a negative response.* Moreover, the up U and down D designations are interchanged. The TMP for a given n_o is based on Eq. 9-11 as before.

The reader will note that the OSTR strategy is actually the Langlie routine applied to a transformed response curve. Moreover, the usual or "standard" Langlie procedure, which may be described by taking Wetherill's $n_o = 1$, may be used to estimate the median or 50% point of the transformed response curve. The solution of Eq. 9-11 in terms of $F(x)$ for the value of $P = 0.50$ gives the probability value of the original response function corresponding to the 50% point of the transformed response, and this is the TMP. Note by observing Table 9-2 that for $n_o = 3$, for example, the value referred to is 0.7937 in the upper tail and 0.2063 in the lower tail.

Finally, a point of some interest concerning the design of an optimum strategy is that the test procedure must close to finer and finer intervals about the desired percentage point, or desired quantile or percentile. However, a fixed interval of testing, such as the up and down strategy, does not do this. Also if at all possible, it certainly would pay to design the testing strategy so that the analysis of results is made as easy as possible.

*For the lower percentage points, the transformation is $(1 - q^{n_o})$ instead of p^{n_o}.

TABLE 9-2

CHARACTERISTICS OF SOME TRANSFORMED RESPONSE STRATEGIES (Ref. 3)

n_o	Response Type* D if $p>0.5$, U if $p<0.5$	Response Type* U if $p>0.5$, D if $p<0.5$	Transformation, $P=$	Percentage Point Estimated $p<0.5$	Percentage Point Estimated $p>0.5$
2	XX	X0, 0	p^2**	0.2929	0.7071
3	XXX	XX0, X0, X	p^3	0.2063	0.7937
3	XXX, XX0X	XX00, X0, 0	$p^3(2-p)$	0.2664	0.7336
4	XXXX	XXX0, XX0, X0, 0	p^4	0.1591	0.8409
4	XXXX, XXX0X	XXX00, XX0, X0, 0	$p^4(2-p)$	0.1959	0.8041
5	XXXXX	XXXX0, XXX0, XX0 X0, 0	p^5	0.12945	0.87055
5	XXXXX, XXXX0X	XXXX00, XXXX0 XX0, X0, 0	$p^5(2-p)$	0.1540	0.8460
6	XXXXXX	XXXXX0, etc.	p^6	0.1092	0.8908
7	XXXXXXX	XXXXXX0, etc.	p^7	0.0944	0.9056
8	XXXXXXXX	XXXXXXX0, etc.	p^8	0.0829	0.9171
9	XXXXXXXXX	XXXXXXXX0, etc.	p^9	0.0740	0.9260
10	XXXXXXXXXX	XXXXXXXXX0, etc.	p^{10}	0.0670	0.9330
14	XXXXXXXXXXXXXX	XXXXXXXXXXXXX0, etc.	p^{14}	0.0484	0.9516

*For $p>0.5$, X = response and 0 = nonresponse.
For $p<0.5$, X = nonresponse and 0 = response.
**For the lower percentage points, use $1-q^{n_o}=1-(1-p)^{n_o}$

Thus, for example, a fairly complex strategy, when used along with a rather simple analysis, probably would, on an overall basis, prove to be very acceptable in practice.

We have placed considerable interest in our discussion on strategies involving one "shot" per level of test. The case of several shots per level is considered in the next paragraph on estimation.

9-4 ESTIMATION OF PARAMETERS

Unfortunately, the reader may have noticed from the discussion of test strategies in par. 9-3 that parameter estimation for sensitivity analyses or models is not a straightforward process. In view of its efficiency, the Fisher method of ML is ordinarily used although least squares procedures or other methods of estimation, such as minimum chi-square (MCS) techniques, may also be employed. We will cover ML estimation first for the normal distribution before any discussion of other estimation techniques.

We should remark, however, that graphical procedures may be used, and this is perhaps especially desirable for the case in which one has sensitivity data from a complete rundown test or results from an experiment giving the proportions of positive responses at each level of stimulus. Here, for example, one may use normal probability paper and plot the cumulative fraction of positive responses vs the stimulus level or independent variable to estimate the mean and standard deviation. Another reason for using graphical estimates, at least initially, is that the ML procedures require good starting values and a number of iterations, so that such estimates may prove valuable indeed. Also any past information on rough values of the mean and standard deviation would be quite helpful in the iteration process.

As already indicated, our approach to the estimation problem will be primarily for the nonuniform intervals of testing where only a single response is obtained at each level of stimulus, and usually the experimenter has aimed to secure the minimum number of tests that give some positive and negative responses in the zone of mixed results in which testing has been more or less assured by the strategy.

9-4.1 MAXIMUM LIKELIHOOD ESTIMATION FOR THE NORMAL MODEL

In 1956, with some key Army applications in mind, Golub and Grubbs (Ref. 4), performed a study of ML estimation for the normal model, which was then widely assumed in connection with penetration of armor investigations and the acceptance of lots of armor plate from manufacturers. In their particular approach, the probability of a penetration p_i was taken as the integral of Eq. 9-1 up to the point of the striking velocity x_i, such as indicated numerically in Eq. 9-3. The likelihood of the sample results using the random variable $\delta_i = 1$, 0—which depends on whether a penetration or nonpenetration occurred—was as given in Eq. 9-8. To simplify the algebra a bit further, the logarithm of the sample likelihood of occurrence may be taken to give

$$\ln P = \sum_i [\delta_i \ln p_i + (1 - \delta_i) \ln q_i] \tag{9-12}$$

where

$$q_i = 1 - p_i \tag{9-13}$$

and the p_i and q_i both involve normal integrals that contain the unknown mean μ and standard deviation σ. Then the differentiation of Eq. 9-12 with respect to both μ and σ gives two equations with these unknowns that may be iterated upon by using some technique, such as the Newton-Raphson method, to determine the μ and σ.

In view of a very elegant study by DiDonato and Jarnagin (Ref. 9) relative to convergence properties and estimation procedures for ML estimation for the normal model, we will follow their analysis. The method of DiDonato and Jarnagin (Ref. 9) is to identify the total sample size for the test results as N (instead of n) and to divide the observations into n penetrations and m nonpenetrations, so that

$$n + m = N. \tag{9-14}$$

This means that the logarithm of the likelihood function L of sample results is

$$L = \ln P = \sum_{i=1}^{n} \ln p_i + \sum_{j=1}^{m} \ln q_i. \tag{9-15}$$

The x_i for which there are positive responses or penetrations are labeled as $a_1, a_2, \ldots, a_n$; those x_j for which there are nonresponses or nonpenetrations are labeled $b_1, b_2, \ldots, b_m$.

DiDonato and Jarnagin (Ref. 9) then deal with transformed parameters (to effect linearization), which are determined from

$$\alpha = \mu/\sigma \tag{9-16}$$

$$\beta = 1/\sigma > 0. \tag{9-17}$$

Finally, instead of using the original standardized normal variates

$$z_i = z(x_i) = (x_i - \mu)/\sigma \tag{9-18}$$

new variates defined by s_i and t_j are taken as

$$s_i = a_i\beta - \alpha \tag{9-19}$$

$$t_j = b_j\beta - \alpha. \tag{9-20}$$

Actually, the p_i and q_i are transformed to be represented as

$$p_i = p(s_i) \text{ and } q_j = q(t_j).$$

Next, in accordance with DiDonato and Jarnigan (Ref. 9), we define and determine the following partial derivatives of the logarithm of the likelihood L in Eq. 9-15:

$$L_\alpha = \partial L/\partial\alpha = \sum_{j=1}^{m} v_j/q_j - \sum_{i=1}^{n} u_i/p_i \tag{9-21}$$

$$L_\beta = \partial L/\partial\beta = \sum_{i=1}^{n} a_i(u_i/p_i) - \sum_{j=1}^{m} b_j(v_j/q_j) \tag{9-22}$$

$$L_{\alpha\alpha} = -\sum_{j=1}^{m}(v_j/q_j)(v_j/q_j - t_j) - \sum_{i=1}^{n}(u_i/p_i)(u_i/p_i + s_i) \tag{9-23}$$

$$L_{\alpha\beta} = \sum_{j=1}^{m} b_j(v_j/q_j)(v_j/q_j - t_j) + \sum_{i=1}^{n} a_i(u_i/p_i)(u_i/p_i + s_i) \tag{9-24}$$

and finally

$$L_{\beta\beta} = -\sum_{j=1}^{m} b_j^2(v_j/q_j)(v_j/q_j - t_j) - \sum_{i=1}^{n} a_i^2(u_i/p_i)(u_i/p_i + s_i) \tag{9-25}$$

where

$$u_i = u_i(s_i) = (1/\sqrt{2\pi})\exp(-s_i^2/2) \tag{9-26}$$

$$v_j = v_j(t_j) = (1/\sqrt{2\pi})\exp(-t_j^2/2). \tag{9-27}$$

The five partial derivatives given in Eqs. 9-21 through 9-25 are used in the iteration process to estimate the values of the transformed parameters α and β, which in turn are finally transformed to the values μ and σ by applying Eqs. 9-16 and 9-17.

DiDonato and Jarnagin (Ref. 9) give a very comprehensive analysis of the existence and convergence properties of the estimates of the unknown parameters pointing out in particular the conditions on the a_i and b_j for which the logarithm of the likelihood function L has a unique maximum. Hence these authors give the necessary and sufficient conditions for L to have a maximum at the final iterated values or point (α,β).

The paper of DiDonato and Jarnagin (Ref. 9) is a somewhat condensed version of a more extensive study, and the full mathematical details of their investigations and analyses are covered in Ref. 10. In fact, DiDonato and Jarnagin (Ref. 10) prove that the logarithm of the likelihood function L in Eq. 9-15 attains a unique global maximum for the estimated parameters α and β attained, and they show that their algorithm, which is a modified form of the Newton-Raphson iterative procedure, does guarantee global convergence. The two iterative equations used in the estimation procedure are given by

$$L_{\alpha\alpha}\Delta\alpha + L_{\alpha\beta}\Delta\beta = L_\alpha \tag{9-28}$$

$$L_{\alpha\beta}\Delta\alpha + L_{\beta\beta}\Delta\beta = L_\beta \tag{9-29}$$

where the quantities $\Delta\alpha$ and $\Delta\beta$ are the changes in the old values of α and β, respectively, calculated at each stage of the iteration. Thus one may start with initial estimates α_0 and β_0 and—by substituting the data into the two first partial derivatives on the right-hand side (RHS) of Eqs. 9-28 and 9-29 and into the three second order partial derivatives on the left-hand side (LHS)—he may solve for the $\Delta\alpha$ and $\Delta\beta$. These differences lead to the next values of $\hat{\alpha}$ and $\hat{\beta}$ to use in the partial derivatives, which are

$$\alpha_1 = \alpha_o + \Delta\alpha \tag{9-30}$$

$$\beta_1 = \beta_o + \Delta\beta. \tag{9-31}$$

The process continues in this manner to some stage n, for which there are very insignificant changes in the newest estimates of the parameters α and β, and finally, the estimated mean and standard deviation of the normal model are determined.

DiDonato and Jarnagin (Ref. 9) indicate that their computer program always converges to the proper estimates no matter what the starting values or initial estimates are. They also indicate that the ordinary Newton-Raphson method will converge irrespective of initial estimates too, so it would seem that if one has at hand some suitable "mixed results" for responses and nonresponses, convergence should be of no concern whatever.

A Naval Weapons Laboratory computer program is available in Ref. 10 for determining and plotting 95% and 50% confidence ellipses for the parameters α and β; the details are presented in Ref. 10 also.

To indicate an illustrative application (Example 9-1), we will give some actual data on a penetration-of-armor plate test, which has been used in Ref. 4.

Example 9-1:

In a ballistic test of 90-mm AP projectiles against rolled homogeneous plate, only five striking velocities along with armor plate response were available for the determination of the median or $V_{0.50}$ level of stimulus. They were

Striking Velocity, ft/s	Condition of Impact
2415	Nonpenetration
2415	Nonpenetration
2423	Penetration
2433	Nonpenetration
2453	Penetration.

With these data find the level of striking velocity for which 50% penetrations would occur and the standard deviation of the assumed normal distribution of penetrations and nonpenetrations.

Observe that the original data have been rearranged in increasing order of striking velocity against the armor plate. We note, for example, that although there is bound to be some random scatter in the muzzle velocities of the AP projectiles fired from a gun, it happened that two striking velocities were the same, 2415 ft/s, and neither of the two projectiles penetrated the plate. There was a penetration at 2423 ft/s, nevertheless, and the highest velocity of 2453 ft/s resulted in a penetration. However, the most significant feature of the data is that, although the projectile with 2423 ft/s gave a penetration, we have a higher striking velocity of 2433 ft/s, which resulted in no penetration. Thus we have a "contradiction" or an indication of being within the zone of "mixed" results, which is always desirable. Hence we should have proper data in this test, which would be analyzable in the sensitivity analysis sense. Moreover, we surely have a small sample and can get some idea as to how well the analysis will proceed.

As indicated, we will assume a cumulative normal distribution describing the zone of mixed results going from zero penetrations at the lower velocities to 100% penetrations at some higher striking velocity and will attempt to estimate both the median and the standard deviation. To do this, however, we will need starting estimates of both. For a starting estimate of the $V_{0.50}$, we may take the average of the highest velocity with no penetration and the lowest velocity occurring with a penetration. This means that we take the initial estimate of $V_{0.50}$ to be $(2423 + 2433)/2 = 2428$ ft/s.

For an initial estimate of the standard deviation of the zone of mixed results, some past data for the projectile-plate combination indicated that the point from no penetrations to 100% penetrations might be about 100 ft/s and surely would not be less than 80 ft/s. Hence a standard deviation of 20 ft/s can be taken as the initial σ. This means that for the parameters in the DiDonato-Jarnagin algorithm, we would have

$$\alpha = \mu/\sigma = 2428/20 = 121.4 \text{ and } \beta = 1/\sigma = 1/20 = 0.05.$$

With these *initial* estimates of the DiDonato-Jarnagin parameters, all of the derivatives in Eqs. 9-28 and 9-29 are calculated, the values inserted, and the changes $\Delta\alpha$ and $\Delta\beta$ are computed. From these latter indicated changes, new values of α and β are calculated and the process continued to the desired degree of accuracy. It will be found through iterative computations that

$$\hat{\alpha} = 162.11 \qquad \text{and} \qquad \hat{\beta} = 0.067$$

or

$$\hat{\mu} = 2431.6 \text{ ft/s} \qquad \text{and} \qquad \hat{\sigma} = 15.0 \text{ ft/s}.$$

The DiDonato-Jarnagin computer program for their algorithm is included with this chapter as Computer Program 9-1, Appendix 9A, for interested users. For those investigators who prefer to work directly in terms of the normal population mean μ and σ, the mathematical and statistical details are included in Ref. 4. A computer program for this case is available from the Director, US Army Ballistic Research Laboratory, Aberdeen Proving Ground, MD 21005, which also includes the estimate of the variance-covariance matrix. The variance-covariance matrix is determined to obtain estimates of the asymptotic standard errors of the estimated mean μ and standard deviation σ of the assumed normal distribution of proportions or chances of penetrations. In this connection, we find from p. 265 of Ref. 4 that

$$\sigma_{\hat{\mu}} = 10.7 \text{ ft/s} \qquad \text{and} \qquad \sigma_{\hat{\sigma}} = 12.5 \text{ ft/s}.$$

Thus these results show that the estimated standard error of 10.7 ft/s for the estimated population mean is quite satisfactory, but the estimated standard error of 12.5 ft/s for the estimated standard deviation is nearly as large as the population sigma itself. Perhaps this would indicate that the (up and down) strategy, which was used in this test, along with the rather small sample size, does not lead to a precise estimate of the population σ. In fact, it would probably be found that the sample size for the test would have to be increased enormously to reduce the standard error $\sigma_{\hat{\sigma}}$ of the $\hat{\sigma}$ to approximately 2 or 3 ft/s.

9-4.2 MAXIMUM LIKELIHOOD ESTIMATION FOR THE LOGISTIC DISTRIBUTION

The ML estimation of parameters for the logistic model of Eq. 9-5 proceeds along similar lines to that indicated for the normal distribution in par. 9-4.1. Moreover, as we stated earlier, Wetherill (Ref. 7) indicates that there is little difference to be found in the use of the normal model compared to the logistic model, with the advantage that the logistic model is somewhat easier to deal with analytically or as a computer program for simulation experiments.

If we are dealing with the situation in which only single tests at each of several stimulus levels are available for analysis, the likelihood function for the observed sample may be taken as in Eq. 9-15 with the stipulation that for the logistic model we now use

$$p_i = F(x_i) = \{1 + \exp[-(\alpha + \beta x_i)]\}^{-1} \tag{9-32}$$

as in Eq. 9-5. Furthermore, one may proceed to obtain partial derivatives for the logistic model along lines similar to those indicated for the normal distribution in Eqs. 9-21 through 9-25 and finally use Eqs. 9-28 and 9-29 for the iteration process from which to determine the parameters α and β. We will not record such similar details here but will leave them for any Army investigators who may find possible applications for the logistic

distribution. In addition, we suggest that some investigators will be interested in comparing normal fits to logistic fits of selected data. We will outline the ML estimation of the parameters for the logistic distribution for the case for which there are several items tested at each of a number of levels of stimulus.

Suppose that there are n_i items tested at level x_i and that r_i items respond to that level of stimulation. Now i may be a general number of different levels, $i = 1, 2, 3$, etc., and the estimate of the proportion $\hat{p}$ of responses at any level i is

$$\hat{p} = r_i/n_i. \tag{9-33}$$

The true proportion of positive responses is given by Eq. 9-32. We have not indicated a particular strategy of testing because that is immaterial. The ML estimators $\hat{\alpha}$ of α and $\hat{\beta}$ of β for the logit (logistic) model are obtained from the simultaneous equations

$$\sum_i n_i\hat{p}_i = \sum_i r_i \tag{9-34}$$

$$\sum_i n_i x_i\hat{p}_i = \sum_i r_i x_i \tag{9-35}$$

where

$$\hat{p}_i = \{1 + \exp[-(\hat{\alpha} + \hat{\beta}x_i)]\}^{-1}. \tag{9-36}$$

Thus with the values of n_i, x_i, and r_i substituted into Eqs. 9-34 and 9-35, there are two equations and two unknowns, so that at least theoretically a solution for the parameters α and β is possible. Although the solution may not be so straightforward, it clearly does not involve integrals as does the normal model. Speaking generally, however, the ML estimation of the two parameters for the logistic model does require iterative methods for a solution. For this reason, we will look at another technique for determining α and β.

For the logistic model it is well-known that there is a straight-line transformation for this function, and it is easily found from what is widely referred to as the "logit". In this connection, observe either Eq. 9-32 or Eq. 9-36, which includes estimates of the parameters, and note that the transformation or "logit" of p_i involving the logarithm

$$\text{logit}\, p_i = \ln(p_i/q_i) = \alpha + \beta x_i \tag{9-37}$$

is indeed linear. In view of this and the usual contention that iteration is an undesirable process for many investigators in laboratories who want quick, practical answers, Berkson (Ref. 11) developed a noniterative solution that is called the "minimum logit χ^2 estimate". This is defined by the minimization of the following quantity called the "logit χ^2", i.e.,

$$\chi^2(\text{logit}) = \sum_i [r_i(n_i - r_i)/n_i]\{\ln[r_i/(n_i - r_i)] - \hat{\alpha} - \hat{\beta}x_i\}^2. \tag{9-38}$$

The latter two terms in the brace of Eq. 9-38 are the negative of the estimated value of the logit. Berkson (Ref. 11) shows that the normal equations for his least squares fit of the logistic distribution, i.e., the minimum logit χ^2 estimates of α and β, may be found from

$$\sum_i [r_i(n_i - r_i)/n_i]\{\ln[r_i/(n_i - r_i)] - \hat{\alpha} - \hat{\beta}x_i\} = 0 \tag{9-39}$$

and

$$\sum_i [x_i r_i(n_i - r_i)/n_i]\{\ln[r_i/(n_i - r_i)] - \hat{\alpha} - \hat{\beta}x_i\} = 0. \tag{9-40}$$

Note that although the calculations of Eqs. 9-39 and 9-40 may still be considered to be a bit tedious, they can nevertheless be solved for the parameters without iteration. The quantity on the RHS of Eq. 9-38 is asymptotically distributed as Pearson's chi-square, and as pointed out by Berkson (Ref. 11), it has the same asymptotic properties as Fisher's ML method does in terms of the minimum logit χ^2 estimates of Eqs. 9-39 and 9-40. In actual computations for the summing process of these two equations, one should take note that observed responses of $r_i = 0$ or $r_i = n_i$ are not to be included. Thus only those steps with observed proportions of responses between zero and unity need be included in calculations because the others do not add any relevant information, or at least very little weight to the overall analysis although there can be much disagreement on the matter. For example, Berkson (Ref. 11) includes an Appendix Note 3 in his paper on the cases of zero and 100% responses or "survivors" in his terminology. He suggests for the case of $r_i = 0$ responses that the working value of $\hat{p}_i = 1/(2n_i)$ should be used, and for the case of $r_i = n_i$ a corresponding observed $\hat{p}_i = 1 - 1/(2n_i)$ should be used. Berkson's arguments seem to be based on the realistic viewpoint that all of the data contain some information of value and, hence, should be used. Although we will not use Berkson's recommendations in our illustrative Example 9-2, interested readers will want to study his paper (Ref. 11).

Example 9-2:

A new artillery primer was developed to be more sensitive than the standard artillery primer, which was considered too difficult to initiate and, in fact, previously had given too high a percentage of duds. In primer sensitivity drop tests using a 2-lb ball to drop on the firing pin, the average drop height for the standard primer distribution was found to be 15 in. Fifty of the new primers were tested at each drop height of 8 in. to 18 in. at spacings of 2 in. in a complete rundown test; the data giving the numbers of responses or proper functions of the primers are listed in Table 9-3.

Is there any evidence that the newly developed artillery primer is more sensitive to initiation than the old standard primer? We should assume in this connection that the flame properties for initiating the propellant are satisfactory.

TABLE 9-3

RESULTS OF PRIMER SENSITIVITY DROP TEST

Height of Drop x_i, in.	Number Tested n_i	Number Functions r_i
8	50	0
10	50	11
12	50	19
14	50	32
16	50	38
18	50	50

The reader may verify, by substituting into Eqs. 9-39 and 9-40 and summing as indicated, that one arrives at the following two equations for estimating α and β:

$$41.000\hat{\alpha} + 534.36\hat{\beta} = 0.5143$$

$$534.360\hat{\alpha} + 7146.96\hat{\beta} = 83.1970.$$

As indicated earlier, we have not included the endpoint estimated fractions of responses of 0/50 at 8 in. and that of 50/50 at 18 in. in the calculations. Solution of these two numerical equations establishes that

$$\hat{\alpha} = -5.449$$

and

$$\hat{\beta} = 0.419.$$

The estimated mean of the distribution from Eq. 9-6 is

$$E(x) = -(-5.449)/0.419 = 13.00 \text{ in.}$$

Since the older standard primer exhibited a mean drop height for functioning equal to 15 in., we conclude that the newly developed primer is more sensitive and, hence, should produce fewer duds.

It can be shown for the logistic distribution of Eq. 9-5 that the standard deviation of the variable x is

$$\sigma(x) = \pi/(\sqrt{3}\,\beta) \tag{9-41}$$

so that for this example, we have

$$\hat{\sigma}(x) = 4.33 \text{ in.}$$

which seems to be a very reasonable value judging from the data that the distribution seems to be about 10 in. or 2.31 sigmas wide.

The reader may like to fit a normal distribution to the data of Table 9-3 by either the ML estimation process or any other selected method and make a comparison with the logistic fit we have obtained by using Berkson's minimum logit chi-square technique. A comparison of the two fitted distributions may be made either by comparing their means and standard deviations or by judging agreement between cumulative distributions computed for several levels of stimulus, i.e., heights of drop.

Perhaps it is of some interest to record that if one knows the standard deviation of x very accurately, i.e., the sigma on the original scale, which from Eq. 9-41 depends on one parameter β, Berkson (Ref. 11) states that the parameter α of Eq. 9-5 may be found (by using an explicit expression of Dr. William Taylor) from

$$\hat{\alpha} = 0.5 \ln\{\Sigma \hat{p}_i^2 \exp(-\beta x_i)/[\Sigma \hat{q}_i^2 \exp(\beta x_i)]\}. \tag{9-42}$$

Moreover, if this last calculation were divided by the known β and the sign changed, it would give the mean or 50% point.

Following up on an earlier remark that there seems to be little choice between the use of the normal distribution and the logistic distribution in sensitivity analysis studies since the logistic is more tractable to handle analytically, it is now of some interest to comment on the use of the ML estimation compared to the minimum chi-square analysis procedure. In 1974 Little (Ref. 12) made a mean square error comparison associated with median response estimation for the normal and logistic distributions, and he included both the ML and MCS techniques. Thus in his simulation analysis Little (Ref. 12) really had four estimates for comparisons in terms of their mean square errors. He found "in broad perspective" that there is little difference among the mean square errors for these four estimators regardless of sample size or stimulus level spacing. Little assumed for the median estimate study that the standard deviations of both distributions were unity, so that the most general type of study was not made. However, in his study he found for "wide" spacing of the stimulus level, i.e., the ratio of spacing to the standard deviation is about 1.5, the mean square errors for ML and MCS were identical for all practical purposes. But for either the "recommended" spacing of about 1σ or for "narrow" spacing of about 0.667σ, the mean square error of the MCS estimation procedure was smaller than that for the ML technique when the stimulus level was within about 1.5σ of the true mean. Otherwise, the ML estimation provides a smaller mean square error than does MCS, and it is more uniform or stable.

Little (Ref. 12) also was able to compare the normal and logistic distributions somewhat. He found that for the initial stimulus level within approximately 2.5σ from the true mean, the logistic distribution would provide a smaller mean square error than would the normal distribution. The normal distribution would provide smaller mean square errors only for the very small sample sizes and the narrower spacings when the initial stimulus level has deviated substantially from the true mean or median of the population.

9-4.3 MAXIMUM LIKELIHOOD ESTIMATION FOR THE WEIBULL MODEL

Although historically the normal and the logistic models have been applied very widely to the analysis of bioassay-type data, and also to many Army investigations, there could always be the criticism that they are not general enough or sufficiently "robust" to describe accurately many important applications. Thus a criticism of the normal model is that it is a two-parameter symmetric distribution and, hence, should not be used to represent skew data. On the other hand, the Weibull model can be used to represent almost any shape. (See, for example, the curves of Fig. 4 of Ref. 3 or those of Fig. 21-7 of the *Army Weapon Systems Analysis* Handbook, Ref. 13.) This statement applies to either the two-parameter or the three-parameter Weibull distribution, i.e., whether $\gamma = 0$ or not in Eq. 9-43.

Generally speaking, the ML estimation of the parameters for the Weibull distribution in sensitivity analysis proceeds as for the normal and logistic models. Thus if, as in Eq. 9-12, we take the logarithm of the general likelihood probability of the sample, which is the L of Eq. 9-15, and find partial derivatives with respect to each of the parameters, which are equated to zero, we have a set of as many equations as there are unknown parameters. These may be solved, especially by computers, for the estimates of the unknown parameters. Again, the δ_i are taken as either unity or zero, depending on whether there is a response or not, although now the p_i is the Weibull form

$$p_i = F(x_i) = 1 - \exp\{-[(x_i - \gamma)/\theta]^{\beta}\} \tag{9-43}$$

where

$$\theta = \sigma^{1/\beta} \tag{9-44}$$

which is the form used by Einbinder (Ref. 3) except that he also uses the shape parameter α instead of our β. In statistical analyses it is the location parameter γ that is troublesome because it is the absolute start of nonzero frequencies. However, some of the difficulty may be avoided by taking different values of the location parameter γ and subtracting the assumed parameter values from the stimulus levels while simultaneously trying to determine the best fit to the observed data by the proper choice of γ.

In view of the likelihood of more and more applications of the Weibull model in future investigations, we will outline the mathematical and statistical details for establishing the iterative equations only for the two-parameter Weibull distribution and otherwise recommend Einbinder's computer program as indicated in Ref. 3, which is included here as Computer Program 9-2, Appendix 9B. To sketch the types of analytical functions and techniques of iteration for the two-parameter Weibull model, with the ML estimation approach similar to that of the normal and logistic models, we will define

$$y_i = x_i - \gamma \tag{9-45}$$

and hence use the form

$$p_i = F(y_i) = 1 - \exp(-y_i^{\beta}/\sigma) \tag{9-46}$$

along with

$$q_i = 1 - p_i. \tag{9-47}$$

Then with this notation we may, by reference to Eq. 9-15, see that

$$L_{\beta} = \sum_{i=1}^{n} q_i y_i^{\beta} (\ln y_i)/(\sigma p_i) - \sum_{j=1}^{m} y_j^{\beta} (\ln y_i)/\sigma \tag{9-48}$$

$$L_{\sigma} = -\sum_{i=1}^{n} q_i y_i^{\beta}/(\sigma^2 p_i) + \sum_{j=1}^{m} y_j^{\beta}/\sigma^2 \tag{9-49}$$

$$L_{\beta\beta} = \sum_{i=1}^{n} [q_i y_i^{\beta} (\ln y_i)^2 (1 - y_i^{\beta}/\sigma)/(\sigma p_i) - q_i^2 y_i^{2\beta} (\ln y_i)^2/(\sigma^2 p_i^2)] - \sum_{j=1}^{m} y_i^{\beta} (\ln y_i)^2/\sigma \tag{9-50}$$

$$L_{\sigma\sigma} = \sum_{i=1}^{n} [2q_i y_i^{\beta}/(\sigma^3 p_i) - q_i^2 y_i^{2\beta}/(\sigma^4 p_i) - q_i y_i^{2\beta}/(\sigma^4 p_i)] + 2 \sum_{j=1}^{m} y_i^{\beta}/\sigma^3 \tag{9-51}$$

$$L_{\beta\sigma} = \sum_{i=1}^{n} (q_i y_i^{2\beta} - \sigma q_i y_i^{\beta} - q_i^2 y_i^{2\beta})(\ln y_i)/(\sigma^3 p_i^2) - \sum_{j=1}^{m} y_i^{\beta} (\ln y_i)/\sigma^2. \tag{9-52}$$

Recall that we are seeking estimates of the shape parameter β and the scale parameter σ, and theoretically at least we could equate Eqs. 9-48 and 9-49 to zero and solve for these parameters. However, both p_i and q_i involve the unknown parameters so that iterative equations very similar to Eqs. 9-28 and 9-29 for the normal model ordinarily would be used in the solution. Thus we would use

$$L_{\beta\beta}\Delta\beta + L_{\beta\sigma}\Delta\sigma = L_{\beta} \tag{9-53}$$

and

$$L_{\beta\sigma}\Delta\beta + L_{\sigma\sigma}\Delta\sigma = L_{\sigma}. \tag{9-54}$$

In Ref. 3 Einbinder indicates that starting values of the two parameters for the iterative solution may be found by matching two percentage points for a fixed value of the location parameter γ. The Wetherill estimator $\bar{w}$ of par. 9-3.5 may be useful for determining such percentage points. According to Einbinder (Ref. 3), "Convergence problems were encountered in solving the nonlinear equations. A transformation of the data into exponential form based upon the initial estimate of the Weibull parameters was found to stabilize and speed convergence to a solution.". Thus as of this time, convergence properties for the Weibull model have not been fully investigated as DiDonato and Jarnagin (Ref. 9) did for the normal model.

Note that the partial derivatives with respect to the parameters for the Weibull model are really quite involved, and surely a computer is required. However, it should be pointed out that the application of the Weibull model results in much, much more generality since, for a wide variety of shapes of quantal response data, the Weibull model could be fitted much better than either the normal or the logistic model.

As indicated in Refs. 3 and 14, Einbinder has developed a computer program (FORTRAN IV) for the Weibull three-parameter and two-parameter models in connection with the estimation of appropriate parameters for quantal response type data. Einbinder's computer program is included here as Computer Program 9-2, Appendix 9B. (Card decks are available from the Systems Effectiveness/Systems Analysis Branch, US Army Armament Research and Development Command, Large Caliber Weapon System Laboratory, Dover, NJ 07801.) Readers will want to use these computer aids for the analysis of Army quantal response data as needed.

Einbinder also established expressions for the asymptotic variances and covariances of the estimated parameters, and these are included in his computer programs (Refs. 3 and 14). We will illustrate in Example 9-3 his example of Ref. 14 for the fitting of a Weibull model to quantal response data that were taken to develop information on safety distances concerning the detonation probabilities of one high explosive projectile from another in case of an accident setting off one of the projectiles.

Example 9-3:

To study the sensitivity of high explosive projectiles to detonation if a nearby projectile were accidentally initiated, it was desired to conduct a sensitivity analysis type of evaluation. Also since there seemed to be little available information and no theory for the new type of high explosive used, one could not be very positive about the shape and width of the quantal response distribution. In view of this, it appeared desirable not to use the cumulative normal distribution of the fraction of responses to describe the results, but rather to use the

Weibull model for such new data. A schematic sketch of the test situation is shown on Table 9-4 along with the strategy for determining the next stimulus level, and the results from testing. In this connection, a donor round is initiated high order, and the effect on the receptor round insofar as whether or not a detonation occurs is observed. A type 1 or positive response is defined as nondetonation of the receptor by the donor round since an increase in the separation distance results in an increase in the probability of a nondetonation. On the other hand, a detonation would be denoted by a negative or type 0 response.

The specific purpose of the test procedure in this application was to seek out an upper tail area of the safe-separation distance for the two projectiles. It was decided to use $n_0 = 4$ for a Wetherill upper tail-type strategy for testing, and hence the TMP for this particular strategy is 84% (see Table 9-2). Thus from the strategy of testing it would be expected that stimulus levels chosen with corresponding results would lie in this region of the true unknown response distribution. Moreover, a type D or down response requires a decrease in the separation distance only if four consecutive responses occur in nondetonations at a level, and this would be described by the result or series 1111 in four consecutive trials. The occurrence of a detonation, i.e., a zero response, for any round prior to the fourth 1 or nondetonation would result in a type U or up outcome. The lower limit at which detonation would occur would be taken as no separation distance, i.e., $A = 0$, and the separation distance for which no detonations would ever be expected to occur was estimated to be $B = 64$ ft. Thus, according to the Langlie test strategy, the first trial was started for a stimulus level equal to the midpoint or 32 ft. The results are indicated on Table 9-4, and it is the aim of the analysis of results to estimate the 84% point of the response distribution and also the shape and scale parameters of the fitted Weibull distribution.

At a 32-ft separation distance the first three test results were nondetonations, but the fourth outcome was a detonation, which indicates that the separation distance must be increased for the fifth shot. Hence, according to the Langlie strategy, one must take the average of the current stimulus level, 32 ft, and the upper boundary $B = 64$ to obtain a separation distance of 48 ft for the next series of tests. At 48 ft the four shots all resulted in nondetonations; therefore, a type D response occurred at the 8th trial—and this brings about the first change number—so that the separation distance must be decreased somewhat now. For the 9th shot we average the last test level of 48 ft with the previous type U response level of 32 and get a 40-ft separation distance for the 9th shot. At trial 12 the signal for another down, or D response, occurred, so that one should include the lower limit A with the 12th separation distance to average for the 13th shot since an equal number of D's and U's could not be found in going from stimulus level 12 to stimulus level 1. For trial number 17 the separation distance is taken as the average of the 16th and the 8th trials since there are two U's and two D's in going from the 16th back to the 8th trial. Finally, using six changes of response type as a stopping rule, all testing was terminated after the 31st trial. Moreover, the criteria for a good zone of mixed results were also satisfied since $x$min1 is less in distance than $x$max0, and $x$min0 is less than $x$max1.

Since we used the Wetherill $n_0 = 4$ and have met the stopping criteria satisfactorily, in summary the Wetherill $\bar{w}$ is approximately 32 ft, which we would take as the 84% point of the safe-separation distance for nondetonations. That is, we would estimate that at approximately 32 ft the chance of an initiation would be about 16%. One notes finally that our choices of the boundaries of zero and 64 ft also seem reasonable.

In the strategy to estimate the 84% point of the cumulative distribution, we did not really assume a particular distribution; therefore, we should concentrate now on fitting, for example, a Weibull model. This will be done for the two-parameter Weibull form by making computations for several assumptions of the location parameter, as indicated in Eq. 9-45 while employing Einbinder's computer program (Ref. 14), which is included as Appendix 9B.

Appendix 9B is for both the fitting of the two-parameter and the three-parameter Weibull models; however, the fitting of the three-parameter Weibull distribution is not so straightforward. In fact, for the iteration processes some further study of the use of initial estimates of the parameters and the convergence properties of the iterations probably is required. At the present time, the use of the normal distribution, as thoroughly investigated by DiDonato and Jarnagin (Ref. 9) and the logistic model, investigated by Wetherill (Ref. 7), may be on more solid ground. Einbinder (Refs. 3 and 14) indicates that for the Weibull model the likelihood function appears flat in the direction of the location parameter, but no real convergence problems have been encountered.

TABLE 9-4

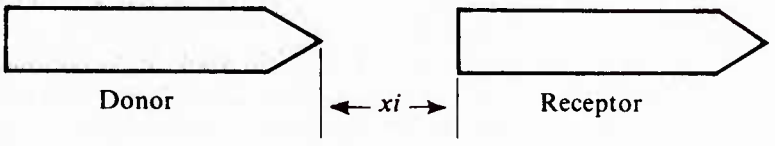
SENSITIVITY ANALYSIS OF PROJECTILE INITIATION

i	xi	Trial Outcome	Response Type	Change Number	Remarks
1	32	1			$xi = (A + B)/2$
2	32	1			Repeat 1
3	32	1			Repeat 1
4	32	0	U		Go up
5	48	1			$x5 = (x4 + B)/2$
6	48	1			Repeat 5
7	48	1			Repeat 5
8	48	1	D	1	Go down
9	40	1			$x9 = (x8 + x4)/2$
10	40	1			Repeat 9
11	40	1			Repeat 9
12	40	1	D		Go down
13	20	0	U	2	$x13 = (x12 + A)/2$
14	30	1			$x14 = (x13 + x12)/2$
15	30	1			Repeat 14
16	30	0	U		Go up
17	39	1			$x17 = (x16 + x8)/2$
18	39	1			Repeat 17
19	39	1			Repeat 17
20	39	1	D	3	Go down
21	34.50	1			$x21 = (x20 + x16)/2$
22	34.50	1			Repeat 21
23	34.50	1			Repeat 21
24	34.50	1	D		Go down
25	27.25	0	U	4	$x25 = (x24 + x13)/2$
26	30.88	1			$x26 = (x25 + x24)/2$
27	30.88	1			Repeat 26
28	30.88	1			Repeat 26
29	30.88	1	D	5	Go down
30	29.06	1			$x30 = (x29 + x25)/2$
31	29.06	0	U	6	Test stopped

NOTES:
1. D = 1111, i.e., no initiations on 4 consecutive trials at same distance. U occurs if initiation (0) occurs before 1111 result.
2. Number of changes of response type equals 6 on trial 31.
3. Min distance with type 1 response is 29.06 on trial 30—*x*min1.
4. Max distance with type 0 response is 32 on trial 4—*x*max0.
5. Test pattern is satisfactory since *x*min1 is less than *x*max0.
6. Criteria for stopping the test are satisfied at trial 31.

A = lower limit = 0
B = upper limit = 64 ft
$x_i = xi$ = separation distance for ith trial; i = trial number

1 = no detonation
0 = detonation



For the data of Example 9-3 Einbinder (Ref. 14) fitted four two-parameter Weibull distributions by assuming that the location parameter γ took on any one of four values $\gamma = 0$, 10, 20, or 25 ft. For each of these four assumed sizes of the location parameter and initial estimates of the shape and scale parameters, the computer program, Appendix 9B, was used to determine the iterated values of the final shape and scale parameters along with the natural logarithm of the likelihood, i.e., a quantity similar to Eq. 9-12 or Eq. 9-15, and the generalized variance. As is well-known, the generalized variance is the determinant of the asymptotic variance-covariance matrix and is quite useful in judging how well the information in the sample is being used to estimate the parameters of the distribution being fitted. This does not, however, guarantee maximization of the likelihood.

The asymptotic variance-covariance matrix for the two-parameter Weibull distribution is found with the aid of the final estimated values of the partial derivatives in Eqs. 9-50 through 9-52. In fact, the asymptotic variance-covariance matrix is given by the following indicated inverse of expected values:

$$\begin{bmatrix} -E(L_{\hat{\beta}\hat{\beta}}) & -E(L_{\hat{\beta}\hat{\sigma}}) \\ -E(L_{\hat{\beta}\hat{\sigma}}) & -E(L_{\hat{\sigma}\hat{\sigma}}) \end{bmatrix}^{-1} \tag{9-55}$$

Hence the generalized variance is the determinant of the inverse matrix given by Eq. 9-55. The matrix of the quantities in Eq. 9-55—without taking the inverse—is known as Fisher's "information matrix".

For Example 9-3 the final quantities computed by Einbinder (Ref. 14) for the four assumed values of the location parameter, the estimates of the shape and scale parameters, the estimates of the logarithm of the likelihood, and the generalized variance are brought together in Table 9-5.

TABLE 9-5

WEIBULL PARAMETER ESTIMATES

Location Parameter γ	Shape Parameter β	Scale Parameter σ	Logarithm Likelihood $\ln L$	Generalized Variance
0	12.01	29.87	–4.3997	37.44
10	8.09	19.86	–4.3944	16.56
20	4.19	9.83	–4.3776	3.99
25	2.22	4.77	–4.3424	0.86

We note from Table 9-5 that there exists a drastic change in the parameter estimates with an increase in the value of the assumed location parameter γ. Also the generalized variance decreases sharply and is smallest at the assumed value of $\gamma = 25$. Thus we should certainly conclude that we are no doubt dealing with a three-parameter instead of a two-parameter Weibull model for the best fit. For a Weibull shape parameter of about 2.22 (last line of Table 9-5), this particular fitted distribution is subnormal or somewhat flatter than the normal distribution, and a bit skewed to the right. Hence we would not expect that either the normal or the logistic models would fit as well as the Weibull although the interested reader may try to obtain proper normal or logistic fits to the same data and to examine the resulting generalized variances for comparative purposes.

In Ref. 14 Einbinder gives estimates of certain percentiles or percentage points (10%, 50%, 84%, 90%, 95%, and 99%) for the safe-separation distances using the fitted Weibull model and each of the values assumed for γ. He also gives the estimated standard deviations of each of these percentiles. These quantities are given in Table 9-6, and the standard errors are listed in parentheses just below each estimated percentile. Reference to Table 9-6 indicates very clearly that the minimum variances for the estimated percentiles occur at the TMP points for the test strategy used, which was 84%. Moreover, the farther away percentile estimates are made

TABLE 9-6

WEIBULL SAFE-SEPARATION DISTANCE PERCENTILES (Nondetonation)*

Percentile, %	$\gamma = 0$ ft	$\gamma = 10$ ft	$\gamma = 20$ ft	$\gamma = 25$ ft
10	24.77	25.03	25.74	26.73
	(3.30)	(3.00)	(2.23)	(1.25)
50	28.97	28.98	29.00	29.04
	(1.35)	(1.33)	(1.25)	(1.07)
84	31.42	31.40	31.36	31.26
	(0.97)	(0.97)	(0.97)	(0.98)
90	32.02	32.01	31.99	31.95
	(1.17)	(1.18)	(1.21)	(1.25)
95	32.73	32.74	32.77	32.86
	(1.50)	(1.53)	(1.61)	(1.75)
99	33.92	33.98	34.15	34.50
	(2.16)	(2.26)	(2.51)	(2.99)

*The upper figures are the estimated safe-separation distance percentiles, and the lower ones in parentheses are the standard deviations of the estimates.

from the TMP point, the greater the standard deviations or variances of the estimators. Of some particular interest is the fact that the estimated percentage points seem to be rather insensitive to the values assumed for possible location parameters. The percentiles or percentage points refer to the areas under the distribution curve fitted, and as might be guessed, the upper percentiles would show less variation with the location parameter than the lower percentiles.

With the estimates of the parameters available, the fitted Weibull model may be used to calculate the probability of initiation of an adjacent projectile as a function of the separation distance. In fact, this has been done by Einbinder (Ref. 14) in his Table 4, which we give here as Table 9-7, for the three separation distances of 30, 34, and 38 ft. Again, the figures in parentheses below each entry are the estimated (asymptotic) standard deviations. For a 30-ft separation distance the detonation probabilities are about equal and do not depend markedly on the choice of the location parameter of the Weibull model although for the larger separation distances of 34 ft and 38 ft, the detonation probabilities vary rather widely with choice of γ. The striking conclusion from Table 9-7 is that the standard errors are very large, relatively. In fact, the coefficients of variation, or ratio of sigma to mean level, are very much larger for Table 9-7 detonation probabilities than for the percentage points of Table 9-6. Also one notes a rather sharp change in detonation chances around 30 ft. For example, from Table 9-6 one notes that the detonation probability is about 16% (i.e., the 84% point) for a separation distance of slightly over 31 ft; however, from Table 9-7 the detonation probability is about double or 34% for 30 ft with only a change of separation distance equal to a bit over 1 ft! In this connection, we also see from Table 9-7 that the standard error of detonation probabilities at 30 ft is half the detonation probabilities themselves! Thus it would be interesting to see whether another model would give smaller sigmas.

A joint confidence region for the Weibull parameters may be estimated by making use of the asymptotic normality of the ML estimators, as is well-known (see Ref. 3).

The percentage points of the Weibull sensitivity model are obtained by solving the following equation for the quantity L_p once a given probability level p is specified:

$$p = 1 - \exp\{-[(L_p - \gamma)/\theta]^{\beta}\} \tag{9-56}$$

TABLE 9-7

ESTIMATED DETONATION PROBABILITIES*

Separation Distance, ft	$\gamma = 0$ ft	$\gamma = 10$ ft	$\gamma = 20$ ft	$\gamma = 25$ ft
30	0.349 (0.145)	0.346 (0.146)	0.341 (0.146)	0.329 (0.145)
34	0.009 (0.032)	0.010 (0.034)	0.012 (0.039)	0.017 (0.044)
38	0.015×10^{-6} (0.41×10^{-6})	0.100×10^{-6} (2.4×10^{-6})	3.350×10^{-6} (55.1×10^{-6})	96.80×10^{-6} (954.0×10^{-6})

*Upper figures are the estimated probabilities of initiation, and the lower ones in parentheses are the asymptotic sigmas.

and if we put

$$q = 1 - p \tag{9-57}$$

we have

$$L_p = \theta(-\ln q)^{1/\beta} + \gamma. \tag{9-58}$$

Asymptotic variances of the estimates of L_p are given in Ref. 3. Thus probabilities for given L_p, or percentage points L_p for given probabilities p, may be determined by using Eq. 9-56 or Eq. 9-58, and asymptotic variances may be found by using well-known statistical approaches.

Einbinder's program, Appendix 9B, is used to calculate parameters and statistical estimates for the reflected Weibull distribution. The cumulative reflected Weibull model is defined as

$$\begin{aligned} F(x) &= \exp\{-[(\gamma_R - x)/\theta]^{\beta}\}, && x \leq \gamma_R \\ &= 1, && \text{otherwise} \end{aligned} \tag{9-59}$$

where

γ_R = starting frequency point for the reflected Weibull model.

The fitting of a reflected Weibull model to a set of observed data is accomplished by reflecting the stress levels and the outcomes about an arbitrary point A. Thus the data for such a case may be transformed to the standard Weibull form by the equations:

$$x_s = 2A - x_i \tag{9-60}$$

$$y_s = 1 - y(x_s) \tag{9-61}$$

where

x_s = transformed stress
y_s = transformed response.

(The shape and scale parameters are invariant under this transformation.)

9-5 SOME REMARKS ON ALLIED WORK

As we have stated earlier, our primary purpose in this chapter is to report on sensitivity analysis work that will likely have Army applications. Moreover, it is for this very reason that we have concentrated on the case for which there is only a single observation for each level of stimulus, no matter whether a uniform spacing of the stimulus levels exists or values were finally arrived at by using nonuniform spacing in the sensitivity experiment. By 1982 the Army has had some 28 design of experiments conferences and some 21 operations research symposia at which a variety of subjects have been presented and discussed, including the topic of sensitivity analysis or quantal response type evaluations. In fact, there has been a wide variety of applications to a number of Army problems—e.g., ballistic limit of armor plate, explosive sensitivity, primer sensitivity, safety distances for storage of ammunition, bioassay in medical or related fields, and rocket motor rupture problems. Thus we will make reference to a few applications and some studies of possible interest to Army investigators.

In connection with sensitivity testing for launch vehicle applications, Gayle (Ref. 15) reported on a computer simulation study of the Bruceton or up and down technique and the probit method, which has been used historically in much of the bioassay analyses. The probit method, or more accurately, the probit transformation, has been widely used to linearize the data when it is assumed that the sensitivity test results follow a cumulative normal distribution. This is done by dealing with standard units of the original data and adding a (large) constant, usually taken as 5 to the number of standard units. Thus in terms of the original data expressed as x units of stimulus, we first have the standardized normal deviates z, i.e.,

$$z = (x - \mu)/\sigma. \tag{9-62}$$

Then if we put

$$y = z + 5 \tag{9-63}$$

we have a new variable y, which is a transformation, but one related to the original cumulative normal probabilities. For example, suppose that the cumulative normal probability is $p = 0.16$, then one may calculate that the equivalent $y = +4.0$.

The quantity y is called the probit of the probability p. We note that y is a linear form of a standardized normal variate and in fact,

$$y = \text{probit } p = \alpha + \beta x \tag{9-64}$$

where we identify that

$$\alpha = 5 - \mu/\sigma \tag{9-65}$$

$$\beta = 1/\sigma. \tag{9-66}$$

In summary, therefore, if we plot the probit y against the original stimulus levels x, for normally distributed data we would expect to get a straight line. Eq. 9-66 gives the slope of the probit line, and Eq. 9-65 gives the intercept. Moreover, the estimate of β is a good measure of the heterogeneity of the sensitivity data under investigation: the smaller the value β, the more heterogeneous the data (which means a large sigma), and the larger the quantity β, the smaller the variability or sigma. An advantage of the probit method is that the probability levels may be preselected, but to equate observed probabilities with the theoretical ones of Eq. 9-64, several observations per level must be used, and the larger this number, the better. (Since there are two parameters, at least two levels must be chosen, and for three or more parameters, least squares should be used.)

The computer simulation carried out by Gayle (Ref. 15) was to compare the up and down and the probit techniques insofar as estimation of the true mean was concerned, but Gayle also was quite interested in the effects of nonnormality, which may have been the case for his problem and estimation of the more extreme percentage points of the distribution—thus, this is the reason for sampling known normal distributions rather extensively and comparing results of analysis. Moreover, Gayle gave particular attention to the probable

nonnormal types of distributions he would encounter and included some bimodal distributions, which are likely to be sampled in practice. Although the up and down and the probit methods would not be strictly comparable, the sampling experiment was carried out so that some valid comparisons could be made. In fact, for each sensitivity experiment Gayle generally obtained about 20 responses at each of some five different levels of stimulus, and each sensitivity test was repeated about 50 times; accordingly, the sampling was somewhat extensive. A selected point of strong interest was that the up and down technique would concentrate testing about the mean, whereas any testing levels could be used for the probit test.

As a result of his sampling experiments and analyses, Gayle (Ref. 15) concluded that the "Probit method for the bimodal distribution was extremely sensitive to the particular levels selected for testing with agreement ranging from poor, in some instances, to ridiculous in others.". For the normal distribution both methods gave good estimates of the true mean level, but when one sampled the distributions departing from normality, the estimates "provided only rough indications of the population parameters", and in the case of bimodal populations the estimates were quite unreliable. For the more extreme percentage points the estimates were found to be very unreliable. We see, therefore, that Gayle's conclusions are similar to Wetherill's (Ref. 7) although the probit method, which has been widely used, was brought into consideration by Gayle (Ref. 15) because he wanted to study the effect of the selection of the stimulus level, especially when there could be the concentration of more than a single test at each stimulus level, if desired, instead of the up and down type of testing technique.

No matter what the underlying, unknown distribution is in an application for a sensitivity analysis type of test, the desire to preselect the particular levels of stimulus should be tied in with an optimum or very useful type of testing strategy. Consequently, much effort has been devoted in recent years to the design of improved testing strategies. Some of these investigations have been reported in the *Proceedings of the Army Design of Experiments Conferences* by, for example, Rothman and Zimmerman (Ref. 16), Alexander and Rothman (Ref. 17), and Little (Ref. 18).

Rothman and Zimmerman (Ref. 16) attempt to extend the work of Gayle (Ref. 15) to more complex-type sensitivity experiments and also to bring into consideration the matter of costs. They consider a sensitivity experiment for which there are n stimulus variables and one for which the cost of each test is at least approximately known as a function of any combination of these variables. They also assume that the cost is no different whether the test response is positive or negative (null). The goal of their study was for a given probability α to estimate a specified portion of that $(n - 1)$ dimensional surface on which the chance of a positive response equals α. Their analysis is based on the use of a loss function L, which would be made up conceptually of two terms: (1) the cost of tolerating a specified variance in the estimate of the surface sought and (2) the cost of testing. The overall problem was stated as the desire to find the experimental design of the testing strategy that would minimize the average value of the loss over the portions of the surface of prime interest. Apparently, there have been no subsequent attempts to extend this type of sensitivity analysis procedure.

In Ref. 17 Alexander and Rothman report on a study to extend knowledge on the testing strategies and analyses for the inverse response problem in sensitivity analyses. The inverse problem is the determination of a stimulus or stress level for which the probability of a positive or null type of response is desired, and usually this might well be an extreme percentage point of some hypothesized distribution. Thus if the stated probability level is α, the aim of Alexander and Rothman in Ref. 17 is to find the stress level $x = x_\alpha$ such that the cumulative probability $F(x_\alpha) = \alpha$. Their work assumes, however, a very general type of response function in that $F(x)$ is assumed to be only monotonic nondecreasing; therefore, the design is otherwise distribution free. Their work draws on the attainments of Dixon and Mood (Ref. 1), the Robbins and Monro test strategy (Ref. 6), and the study of Wetherill (Ref. 7) and results in their (Alexander and Rothman) developing two rather complex designs or test strategies for the purpose of using all the previous information in the sensitivity test to determine the next stress level instead of using only the immediately past test results. (The Langlie procedure of Ref. 5 does this in a way.) Alexander and Rothman (Ref. 17) indicate that their recommended designs "give good results with limited sample sizes" for probability levels of 0.05 or 0.95 and "are still useful in many applications" for even probability levels of either 0.02 or 0.98. One design is appropriate when continued testing on a set of discrete test levels is desired until a specified precision in the estimate of x_α is attained. The other design or strategy is appropriate when the sample size is fixed in advance and there are no

restrictions on the test levels. These two designs have been evaluated by Alexander and Rothman with a Monte Carlo or simulation procedure, and as they say, "It is shown that they compare favorably with existing procedures and with a conjectured asymptotic criterion for the distribution-free inverse response problems.". For details readers should study Alexander and Rothman's paper (Ref. 17). Apparently, there has been no follow-up on this work, and it has not yet appeared in the open literature for any extensive application trials.

Little (Ref. 18) has investigated a "two-point" strategy in planning quantal response experiments for ordnance devices. Little recommends a small sample strategy, which hopefully "should prove to be useful in predicting high reliability [or high safety] for ordnance devices". Little's two-point strategy, stated briefly, uses the Bruceton, or up and down, strategy in the first state of testing to generate two nonzero, nonunity probability points along the assumed response distribution curve. Then in the second stage the Little strategy allocates the remaining specimens to two corresponding stimulus levels such that the variance of the point estimate pertaining to the reliability (safety) of interest is minimized. Apparently, it could be said that the first stage is to "feel out" the zone of mixed results for the purpose of "anchoring" the two ends of a line segment as precisely as possible. Or as Little says (Ref. 18), "In essence, the issue is to find the specimen allocations which minimize the variance associated with extrapolation along the fitted response distribution to a point more remote to the median. Optimally, this minimization requires testing certain specific proportions of the available specimens at carefully selected specific stimulus levels.". This particular strategy was developed, according to Little for analogous use in estimating fatigue reliability (Ref. 19). We recall that reliability means the integral of the distribution curve, preferably from a lower percentage point or probability level to infinity, so that a high value—e.g., 95% or 99%—may be achieved as the chance that the item performs reliably or safely.

If we reflect momentarily on the probit method covered earlier in this paragraph, there was an attempt at linearization that was very analogous to the strategy proposed by Little (Ref. 18) for his two-point technique, and it is well-known that if one is fitting a straight line and knows that the correct curve to fit is a straight line, he may as well divide the total available number of observations equally between two points or segment ends as remote as possible. Expressed analytically, the Little strategy proceeds as follows in determining the two points of testing to minimize the variance of reliability prediction. If we use $\hat{}$ to denote estimate of, the fitted linear response model in terms of a probability level p will be given by

$$\hat{Y} = F^{-1}(\hat{p}) = \hat{\alpha} + \hat{\beta}x \tag{9-67}$$

where x refers to the stress level or stimulus and the probability level $p = F(Y)$ is the distribution of interest, i.e., a normal model, logistic, Weibull, etc. As before, the fitted linear response is indicated on the RHS of Eq. 9-67. Thus we see that for any selected stress level x, there will be an estimated value of Y that is convertible to a probability level p through the model or distribution fitted. Moreover, this means that the variance of the fitted or estimated Y may be obtained from the expression

$$\sigma^2(\hat{Y}) = (dy/dp)^2(pq/n) \tag{9-68}$$

where

$q = 1 - p$ = true probability of response
n = number of specimens tested at the stimulus level x.

Hence it is the quantity or variance (Eq. 9-68) that Little minimizes.

Now if we were to select two stress levels—a low one x_1 and a high one x_2—at which r_1 test specimens of n_1 respond at x_1 and r_2 of n_2 respond at x_2, we have estimates of the p_1 and p_2 that are related to the y_1 and y_2 through $p = F(y)$. Furthermore, if we are interested in a particular stress level x_0 for which we desire to know or assure the proper value of reliability, the minimum variance of the corresponding linear response value y_0 may, as shown by Little (Ref. 18), be determined by the appropriate choice of y_1 and y_2 in the expression

$$\sigma^2_{\hat{y}_0} = \{[(y_2 - y_0)/(\sqrt{n_1}\sigma_{\hat{y}_1})] \pm [(y_1 - y_0)/(\sqrt{n_2}\sigma_{\hat{y}_2})]\}^2/[(n_1 + n_2)(y_2 - y_1)^2]. * \tag{9-69}$$

*The plus sign is taken for extrapolation when y_0 is outside the interval (y_1, y_2), and the minus sign, for interpolation.

It is an interesting fact that if one takes the derivatives of Eq. 9-69 with respect to y_1 and y_2 and equates the results to zero, it can be shown that the optimum values of y_1 and y_2 are independent of the value y_0 corresponding to the x_0 in which we are interested! However, the optimum values of y_1 and y_2 along with the n_1 and n_2 (usually equal) have to be computed numerically from the model of interest. This has been done by Little for the normal, logistic, and Gumbel's extreme value distributions (for the smallest observation) and displayed as a table in Ref. 18. We give the results for the normal and logistic distributions in Table 9-8. From Table 9-8 we see, for the assumption of a normal distribution for the quantal response problem, that half of the specimens should be tested at about the 6% probability level, and the other half at the 94% probability level. For the assumption of a logistic distribution, Table 9-8 indicates that half the available items should be tested at the 8% probability level, and the other half at the 92 percentile.

TABLE 9-8

OPTIMUM y AND p VALUES FOR MINIMUM VARIANCE ESTIMATION OF y_0

Distribution	Optimum y's		Optimum p's	
	y_1	y_2	p_1	p_2
Normal	−1.575	+1.575	0.058	0.942
Logistic	−2.399	+2.399	0.083	0.917

The reader will understand that tests should be carried out at stimulus levels in the zone of mixed results, and not at extreme or very low or high probability levels, because a delicate balance should be attained between all the parameters. Thus Eq. 9-69 would indicate, by observing the denominator, that the two test points should be as far apart as possible, but the variances of the two proportions at the two points of test depend on Eq. 9-68 while the choice of the percentile of particular interest y_0 and the division of the total sample come into play. Hence the need exists for a careful examination of Eq. 9-69. In fact, calculations using Eq. 9-69 show that the standard errors in the denominator of Eq. 9-69 will approach zero for very high or very low percentiles, so that the variance of prediction for the point of interest y_0 does indeed get very large. Thus there are unique values of the stress levels x_1 and x_2, which must not be too far apart or too close together, to minimize Eq. 9-69.

Finally, one may want to select a value or level of precision by using Eq. 9-69, and it becomes very clear that the size of the total sample may be quite important especially for an extreme percentile of interest. (For the assumption of the extreme value distribution of Gumbel, Little indicates in Ref. 18 that the two stress levels should be at the 12% and the 92% probability points, indicating the need for testing well into the upper tail of a very skew distribution.)

In spite of all this enlightenment, we cannot escape the hard fact that for practically all problems of application we do not have any very precise ideas as to what the stress levels should be to give, for example, for any normal population about 6% and 94% responses in a proposed test. In fact, even these two percentages are too close to zero and unity to have much direct application to many Army problems. Thus the real difficulty lies in selecting the two stress levels so that we do not obtain all no responses or all responses because this results in a loss of information (large variance of prediction) or in useless testing. Hence for the optimal linear regression we need to have very accurate initial estimates of the intercept α and the slope β, but even this requirement turns out to be impractical. This is the reason that Little (Ref. 18) has suggested his modified procedure called the overall two-point strategy, and for this he recommends some testing using the up and down strategy in the initial stages of the sensitivity test. Actually, Little (Ref. 18) recommends two versions of the two-point strategy, one for "small" samples of "50 specimens or less", and the other for "large" samples of "100 or more" specimens.

For the small-sample procedure, Little suggests: "(1) Conduct the beginning portion of the test program using an up and down strategy, and (2) change over to testing at only two stimulus levels x_1 and x_2 as soon as two finite values of y_1 and y_2 are established by the up and down portion of the test program.". At this point, however, Little (Ref. 18) suggests a third possibility: (3a) allocating the ratio of the number n_1 tested at the lower level to the number n_2 for the higher level directly as the calculated standard error for the level and inversely as the deviation of the point of prediction y_0 from the level y_i, or otherwise (3b) proceeding to treat the test specimens equally at the optimum two probability levels of Table 9-8 if sufficient information is available. These two levels should be updated as the test progresses, and the iterative procedure may be quite worthwhile when the x_1 and x_2 are closely spaced. Little recommends that the up and down portions of the test program should be at equally spaced intervals, i.e., uniform spacing of approximately one sigma each. An example is given by Little in Ref. 18. It would appear that this treatment of the sensitivity analysis by Little may need further study, especially on getting into the second stages of the test strategy, although there could be some Army applications to which the procedure would apply quite well. The real problem appears to be attempting to test near the desired low and high probability levels, and that knowledge in itself would be quite a lot.

For his "large-sample" procedure Little (Ref. 18) depends on the results of testing the small sample to determine more accurately the two levels of test or stimulus for the remainder of the available large sample.

It is hoped that those Army investigators interested in research will extend this direction of sensitivity analysis.

Ross (Ref. 20) discusses the ML estimation of the "12D dose" for the radiation sterilization of canned food, using data from inoculated pack experiments. Thus Ross' problem of Ref. 20 is to assess the effectiveness of ionizing radiation as a means of food preservation. The so-called "12D dose" is obtained and defined in terms of the probability that an individual organism will be killed; obviously, it is desired that this be very high. Therefore, if the cumulative probability of the chance of death at stimulus level x for an individual organism is $F(x)$, the probability that the organism survives is $1 - F(x)$. It is desired to determine the stimulus level x_c such that

$$1 - F(x_c) = 1 \times 10^{-12} \tag{9-70}$$

i.e., that the chance of survival is only one in a trillion—indeed a very low risk! For the case of a can containing n organisms, the chance that all n organisms are killed is

$$[F(x)]^n \approx \exp\{-n[1 - F(x)]\} \tag{9-71}$$

if n is large and the survival chance $1 - F(x)$ is small.

For this particular problem Ross (Ref. 20) has developed computer programs for certain one- and two-parameter distributions to find the critical value of the stimulus level x_c for the "12D dose" by using the inoculated pack experimental data. The distributions for which Ross had computer programs are the one-parameter and two-parameter exponential distributions, the normal, the lognormal, and the Weibull models. He gives an example in Ref. 20 for parameter estimation of all these models for an inoculated pack radiated at $-30°$ C using C. botulinum spores in canned pork.

In a recent paper Hamilton (Ref. 21) reports on a rather extensive Monte Carlo investigation of robust procedures to estimate the median dosage level in binary-response bioassay-type investigations. Generally, his work is for the situation in which several or many items are tested at various dosage levels. Hamilton takes into account the mean square errors of the estimators for a variety of symmetric tolerance distributions, the sensitivity of the estimator to an anomalous response and the possibility that the estimators are incalculable. He includes a discussion of trimmed estimators.

For a fairly comprehensive introductory account of bioassay-type procedures up to about 1975, the reader might be interested in Hubert's lecture notes (Ref. 22). They contain a very readable account of sensitivity analysis procedures, and Hubert includes an extensive bibliography of 133 publications.

9-6 SUMMARY

A large number of important Army applications involve quantal response (all or nothing) type data, and the basic underlying probability distributions that generate such data may take on a wide variety of shapes. In fact, a zone of mixed results exists such that the proportions of responses may vary from zero to 100%. The analyst thus has the job of hypothesizing a reasonably practical distribution and of trying to estimate the parameters of it as precisely as possible. Also there is naturally some rather intense interest in either low or high percentage points, so that efficient strategies of testing are involved. In this chapter we have covered some of the key methodologies that have been developed over the years and that should prove valuable to analysts in their daily work. The normal distribution, the logistic distribution, and the Weibull models have been presented with the more efficient methods of estimation. In addition, we have indicated some of the better strategies of testing in case the experiments can be designed and conducted. Our procedures discussed here are more or less aimed toward the more usual Army application for which there is only one test per level of stimulus. Therefore, unequally spaced data come within the scope of the analyses covered.

Several illustrative problems or applications have been presented to indicate the probable types of uses of sensitivity analysis models.

REFERENCES

1. W. J. Dixon and A. M. Mood, "A Method for Obtaining and Analyzing Sensitivity Data", Journal of the American Statistical Association **43**, 109-26 (1948).
2. K. A. Brownlee, J. H. Hodges, Jr., and M. Rosenblatt, "The Up-and-Down Method in Small Samples", Journal of the American Statistical Association **48**, 262-77 (1953).
3. Seymour K. Einbinder, *One-Shot Sensitivity Test For Extreme Percentage Points*, Proceedings of the Nineteenth Conference on the Design of Experiments in Army Research, Development and Testing, ARO Report 74-1, Army Research Office, 369-86, 1974.
4. Abraham Golub and Frank E. Grubbs, "Analysis of Sensitivity Experiments When the Level of Stimulus Cannot Be Controlled", Journal of the American Statistical Association **51**, 257-65 (1956).
5. H. J. Langlie, *A Reliability Test Method for "One-Shot" Items*, Proceedings of the Eighth Conference on the Design of Experiments in Army Research, Development and Testing, ARO-D Report 63-2, Army Research Office, 145-65, 1963. (See also H. J. Langlie, *A Reliability Test Method for "One-Shot" Items*, Publication No. U-1792, Aeronutronic Div., Ford Motor Co., Newport Beach, CA, 1962.)
6. H. Robbins and S. Monro, "A Stochastic Approximation Method", Annals of Mathematical Statistics **29**, 400-7 (1951).
7. G. B. Wetherill, "Sequential Estimation of Quantal Response Curves", Journal of the Royal Statistical Society, Series B **25**, 1-48 (1963).
8. D. J. Finney, *Probit Analysis,* Cambridge University Press, Cambridge, England, 1952.
9. A. R. DiDonato and M. P. Jarnagin, Jr., "Maximum Likelihood Estimation in Quantal Response Experiments", SIAM Journal of Applied Mathematics **26**, 447-54 (March 1974).
10. A. R. DiDonato and M. P. Jarnagin, Jr., *Use of the Maximum Likelihood Method Under Quantal Responses for Estimating the Parameters of a Normal Distribution and Its Application to an Armor Penetration Problem*, Technical Report TR-2846, US Naval Weapons Laboratory, Dahlgren, VA, 1972.
11. Joseph Berkson, "Maximum Likelihood and Minimum χ^2 Estimates of the Logistic Function", Journal of the American Statistical Association **50**, 130-62 (1955).
12. R. E. Little, "A Mean Square Error Comparison of Certain Median Response Estimates for the Up-and-Down Method With Small Samples", Journal of the American Statistical Association **69**, 202-6 (1974).
13. DARCOM-P 706-101, Engineering Design Handbook, *Army Weapon Systems Analysis Handbook, Part One*, 1977.
14. Seymour K. Einbinder, *A Sequential Sensitivity Test for Extreme Percentage Points and Estimation Using a Weibull Response Model*, 17th US Army Operations Research Symposium, Fort Lee, VA, 6-9 November 1978.

15. John B. Gayle, *Computer Simulations of Bruceton and Probit Methods of Sensitivity Testing*, Proceedings of the Ninth Conference of the Design of Experiments in Army Research, Development and Testing, ARO-D Report 64-2, US Army Research Office, 205-27, December 1964.
16. D. Rothman and J. M. Zimmerman, *The Design of Complex Sensitivity Experiments*, Proceedings of the Tenth Conference on the Design of Experiments in Army Research, Development and Testing, ARO-D Report 65-3, US Army Research Office, 575-93, October 1965.
17. M. J. Alexander and D. Rothman, *Designs and Analyses for Inverse Response Problems in Sensitivity Testing*, Proceedings of the Eleventh Conference on the Design of Experiments in Army Research, Development and Testing, ARO-D Report 66-2, US Army Research Office, 338-65, May 1966.
18. R. E. Little, *Planning Quantal Response Tests for Ordnance Devices: The Two-Point Strategy*, Proceedings of the Twenty-First Conference on the Design of Experiments in Army Research, Development and Testing, ARO Report 76-2, US Army Research Office, 151-65, May 1976.
19. R. E. Little and E. H. Jebe, *Statistical Design of Fatigue Experiments*, Applied Science Publishers, Ltd., London, England, 1975.
20. Edward W. Ross, Jr., *Maximum Likelihood Estimation of 12D for Inoculated Packs*, Proceedings of the Twenty-Second Conference on the Design of Experiments in Army Research, Development and Testing, ARO Report 77-2, US Army Research Office, 317-26, June 1977.
21. Martin A. Hamilton, "Robust Estimates of the ED50", Journal of the American Statistical Association **74**, 344-54 (June 1979).
22. J. J. Hubert, *Introductory Lectures on Bioassay*, Department of Mathematics and Statistics, University of Guelph, Canada, September 1975.

BIBLIOGRAPHY

Herbert C. Batson, *Bio-Assay: The Quantal Response Assay*, Proceedings of the Eighth Conference on the Design of Experiments in Army Research, Development and Testing, ARO-D Report 63-2, US Army Research Office, 551-78, October 1963.

J. D. Church and E. B. Cobb, "A Modified Spearman Estimator of a Location or Scale Parameter", Journal of the American Statistical Association **69**, 207-10 (March 1974).

W. G. Cochran and M. Davis, "The Robbins-Monro Method for Estimating the Median Lethal Dose", Journal of the Royal Statistical Society, Series B **67**, 28-44 (1965).

H. A. Guess and K. S. Crump, "Maximum Likelihood Estimation of Dose-Response Functions Subject to Absolutely Monotonic Constraints", The Annals of Statistics **6**, 101-11 (January 1978).

L. D. Maxim, A. D. Hendrickson, and D. E. Cullen, "Experimental Design for Sensitivity Testing: The Weibull Model", Technometrics **19**, 405-13 (November 1977).

R. R. Sowden, "Bias and Accuracy of Parameter Estimates in a Quantal Response Model", Biometrika **58**, 595-603 (December 1971).

R. Srinivasan and R. M. Wharton, *Sequential Estimation of Quantile Response Curves*, Proceedings of the Twenty-Fourth Conference on the Design of Experiments in Army Research, Development and Testing, ARO Report 79-2, US Army Research Office, 229-40, 1979.

R. K. Tsutakawa, "Design of Experiment for Bioassay", Journal of the American Statistical Association **67**, 584-90 (September 1972).

H. Weiler, "A Significance Test for Simultaneous Quantal and Quantitative Responses", Technometrics **6**, 273-85 (August 1964).

M. E. Wise, "On Normalizing the Incomplete Beta-Function for Fitting to Dose-Response Curves", Biometrika **47**, 173-5 (1960).

APPENDIX 9A
COMPUTER PROGRAM 9-1

(DiDonato and Jarnagin Maximum Likelihood Estimation
of Normal Distribution Parameters)

SUBROUTINE EPPA

PURPOSE. This routine gives maximum likelihood estimates for the mean μ and standard deviation σ of a normal distribution which governs variations in data from experiments in which the response is quantal in every case. Dosage mortality studies and armor penetration analyses are often based on experiments with quantal responses. These responses are associated with the n input values a_i if they are successes and with the m input values b_j if they are failures. (See reference cited below.) At the user's option plots of the confidence ellipses at the 50% and 95% can be obtained as part of the output.

RESTRICTIONS:

1. minimum $a_i <$ maximum b_j

2. $1/m \sum_{j=1}^{m} b_j < 1/n \sum_{i=1}^{n} a_i$ Ref. 1

RESTRICTION. The total number, $n + m$, of different values for a_i and b_j that can be run is limited only by the amount of memory available to user.

ACCURACY. The accuracy of the estimates for μ and σ can be deduced from the print of the iterations. The program is presently set to terminate when the kth iteration satisfies

$$|\Delta(\mu_k/\sigma_k)| < \epsilon_1 |(\mu_k/\sigma_k)|$$

$$\Delta(1/\sigma_k) < \epsilon_2 (1/\sigma_k)$$

where

$$(1/2)\epsilon_2 = \epsilon_1 = 2.5 \times 10^{-4}.$$

REFERENCES:

1. A. R. DiDonato and M. P. Jarnagin, Jr., *Use of the Maximum Likelihood Method Under Quantal Responses for Estimating the Parameters of a Normal Distribution and Its Application to an Armor Penetration Problem*, Technical Report TR-2846, US Naval Weapons Laboratory, Dahlgren, VA, November 1972.
2. *Users Guide for the CDC 6700 Computing System*, NSWC/DL Technical Report TR-3228, US Naval Surface Weapons Center/Dahlgren Laboratory, Dahlgren, VA, December 1974.

CALLING SEQUENCE:

CALL EPPA (IDENT, k, 1, IØP, ALPHO, BETO, FNA, A, FNB, B, Z5) where

IDENT is an array dimensioned at 8 locations. The Hollerwrith character contents of IDENT will be printed on the top line of the OUTPUT. Up to 80 characters are allowed per job.

k is twice n where n is the number of numerically different a_i values. $n \geq 2$.

(cont'd on next page)

APPENDIX 9A (cont'd)

1	is twice m where m is the number of numerically different b_j values. $m \leq 2$.
IØP	If IØP = 0 then user will receive plots of 50% and 90% confidence ellipses with his output. If IØP ≠ 0 no plots will be made.*
ALPHO BETO	Are user supplied starting values α_0, β_0 to the routine. The user can have the routine compute starting values $\alpha_0 = \mu_0/\sigma_0$ and $\beta_0 = 1/\sigma_0$ by setting BETO ≤ 0.
FNA, A	are $k = 2 \cdot n$ dimensioned arrays. FNA(i) specifies the number of A(i) values to be used, i = 1,2, . . . ,n. A($n+1$), A($n+2$), . . . , A($2n$) and FNA($n+1$), FNA($n+2$), . . . , FNA($2n$) is used by EPPA as work space.
FNB, B	are $1 = 2m$ dimensional arrays. FNB(i) specifies the number of B(i) values to be used, i = 1,2, . . . ,m. B($m+1$), B($m+2$), . . . ,B($2m$) and FNB($m+1$), FNB($m+2$), . . . , FNB($2m$) is used by EPPA as work space.
Z5	is an array dimensioned at 201. It is used by the package of plotting subroutines. See Example below.

*Remark—If the user is using the plotting option, i.e., IØP = 0, then 3 of the 4060-IGS subroutines must be called. They are MØDESG, CRTID, and EXITG. (See 2, p. G-13.)

EXAMPLE:

```
Program Sample (output, tape 51 = output, tape 10, etc.)
Dimension FNA(101), FNB(101), A(101), B(101)
Dimension Ident (8), Z5 (201)
    .
    .
    .
Data
    .
    .
    .
call MODESG (Z5,0)
call CRTID (Z5, 20HN35A111GDKLABXXXXXXX)
IØP = 0
call eppa (Ident, NM, MM, IØP, ALPHO, BETO, FNA, A, FNB, B, Z5)

call EXZTG(Z5)
call Exit
End.
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
      SUBROUTINE LCOM(SUM1,SUM2,N,A,FN,CONST,TONES,TTWOS,TTHRES,TFOUR,
     1TFIVE,ZXY,PDZ)
      COMMON/DANDE/ EP1,EP2,LIMIT,NC
      COMMON/ZZZ/BETA,BETA0,ALPHA,QQ
      COMMON/CPNDF/ENDF
      DIMENSION RATION(5)
      DIMENSION A(N),FN(N),ZXY(N),PDZ(N)
      DATA (RATION(K),K=3,5)/.5,.66666666666667,.75/
      DATA SQPI/.39894228040143/
       ENDF=0.
      SUM1=0.0
      SUM2=0.0
      TONES=0.0
      TTWOS=0.0
      TTHRES=0.0
      TFOUR=0.0
      TFIVE=0.0
      DO 13 I=1,N
      SI=A(I)*BETA0-ALPHA
      ZSI=-SI*SI/2.0
   90 FORMAT ( 1H0, 6E22.14 )
       IF ( ZSI.LT.-675.82 )  GO TO 131
      ZXY(I)=SQPI*EXP(ZSI)
  151 CONTINUE
      IF(ABS(SI).GT.8.0)GO TO 8
      IGO=1
      IF (CONST.GE.0.) GO TO 500
      CHECK = PNDF(-SI,0)
      GO TO 501
  500 CHECK = PNDF(SI,0)
  501 QQ=QQ*CHECK**FN(I)
    6 TONE=ZXY(I)/CHECK
      PDZ(I)=TONE
      TONE=FN(I)*TONE
   60 TEMP=A(I)*TONE
   61 SUM1=SUM1+TEMP
      TTWO=TONE*SI
      TTHREE=TONE*PDZ(I)
      GO TO(121,12 ),IGO
  121 SUM2=A(I)*A(I)*(TTWO+CONST*TTHREE)+SUM2
      GO TO 12
    8 CONSTS=CONST*SI
      IF(CONSTS.GE.0.0)GO TO 133
      QQ=0.0
   89 FORMAT ( 1H0, I10 )
      TONE = PNDF(SI,IFIX(CONST-1.))
      PDZ(I)=TONE
      TONE=FN(I)*TONE
      SUM2=SUM2-ENDF*A(I)*A(I)*CONST*FN(I)
      IGO= 2
      GO TO 60
   12 TONES=TONES+TONE
      TTHRES=TTHRES+TTHREE
      TTWOS=TTWOS+TTWO
      TFOUR=TFOUR+TEMP*SI
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
      TFIVE=TFIVE+TTHREE*A(I)
      GO TO 13
  131 CONTINUE
       ZXY(I)=0.
       GO TO 151
  133 PDZ(I)=0.0
   13 CONTINUE
      RETURN
      END
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
      FUNCTION PNDF(X,IQ)
      COMMON/CPNDF/ENDF
      LOGICAL L1,L2
      DIMENSION A(5),B(5),AA(9),BB(9),AAA(6),BBB(6)
      DATA (A(I),I=1,5) /.1857777061846OE-0,.31611237438706E+1,
     2.11386415415105E+3,.37748523768530E+3,.32093775891385E+4/
      DATA (B(I),I=1,5) /.1                E+1,.23601290952344E+2,
     2.24402463793444E+3,.12826165260774E+4,.28442368334392E+4/
      DATA (AA(I),I=1,9) /.2153115354744OE-7,.56418849698867E+0,
     2.88831497943884E+1,.66119190637142E+2,.29863513819740E+3,
     3.88195222124177E+3,.17120476126341E+4,.20510783778261E+4,
     4.12303393547980E+4/
      DATA (BB(I),I=1,9) /.1                 E+1,.15744926110710E+2,
     2.11769395089131E+3,.53718110186201E+3,.16213895745667E+4,
     3.32907992357335E+4,.43626190901432E+4,.34393676741437E+4,
     4.12303393548037E+4/
      DATA (AAA(I),I=1,6) /-.16315387137302E-1,-.30532663496123E-0,
     2-.36034489994980E-0,-.12578172611123E-0,-.16083785148742E-1,
     3-.65874916152984E-3/
      DATA (BBB(I),I=1,6) /.1                E+1,.25685201922898E+1,
     2.18729528499235E+1,.52790510295143E+0,.60518341312441E-1,
     3.23352049762687E-2/
      DATA C0,C1,C2,C3 /0.,1.,2.,.5/
      DATA C4,C5,C6 /1.4142135623731,.56418958354776,1.7724538509055/
       XSAV=X
      XA = ABS(X)
      Y = X/C4
      YA = ABS(Y)
      S = Y*Y
      PA = C0
      PB = C0
      IF (YA.GT.C3) GO TO 20
      DO 10 I=1,5
      PA = PA*S+A(I)
   10 PB = PB*S+B(I)
      T = (PA/PB)*Y/C2
      IF (IQ.NE.0) GO TO 15
      PNDF = T+C3
      RETURN
   15 PNDF = C3-T
      RETURN
   20 L1 = X.GT.C0.AND.IQ.EQ.0.OR.X.LT.C0.AND.IQ.NE.0
      IF (YA.GE.4.) GO TO 40
      DO 30 I=1,9
      PA = PA*YA+AA(I)
   30 PB = PB*YA+BB(I)
      T = PA/PB
      GO TO 60
   40 L2 = XA.GT.8.
      IF (L1.AND.L2) GO TO 70
      Y = C1/S
      DO 50 I=1,6
      PA = PA*Y+AAA(I)
   50 PB = PB*Y+BBB(I)
      X = PA/PB
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
      T = X*Y
      IF (L2) GO TO 80
      T = (T+C5)/YA
   60 PNDF = EXP(-S)*T/C2
      IF (L1) PNDF = C1-PNDF
       X=XSAV
      RETURN
   70 PNDF = C1
      RETURN
   80 Y = T*C6+C1
      PNDF = XA/Y
      ENDF = X*C6*C2/(Y*Y)
       X=XSAV
      RETURN
      END
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
      SUBROUTINE EPPA ( IDENT,NM,MM, IOP ,ALPHA3,BETA3,FNA,A,FNB,B,Z5 )
C  EPPA - EXPLORATORY PROGRAM FOR PROBIT ANALYSIS
      COMMON/ZZZ/BETA,BETA0,ALPHA,QQ
      COMMON/DANDE/ EP1,EP2,LIMIT,NC
       DIMENSION A(NM),FNA(NM),B(MM),FNB(MM)
       DIMENSION IX(1001),IY(1001),JX(502),JY(502)
       DIMENSION IDENT(8)
      DIMENSION G(100),F(100)
       DIMENSION D1(2),PLT(2),QQSAV(100)
      DIMENSION DLAB(10),CHM(5),CHS(5)
     1         ,ITX(50),ITY(50),XLAB1(50),XLAB2(10),YLAB1(50),YLAB2(10),
     2          IXL(10),IYL(10)
       DIMENSION Z5(200),T652(10)
       DIMENSION T653(11),T655(11)
       DIMENSION T657(14),T659(11),T660(12)
       DIMENSION T661(3) ,T658(5)
       DIMENSION T605(4),T604(2),T606(3)
       DIMENSION T599(10 )
       DIMENSION T662(4),T663(4)
       DIMENSION TEMP(21)
      EQUIVALENCE(ALPHA0,ALPHA,A1),(BETA0,B1)
      EQUIVALENCE(R,AUU),(S,AUS),(T,ASS)
      REAL MU0
      DATA EP1,EP2,LIMIT,NC /2.5E-4,5.E-4,100, 4/
       ALPHA0=ALPHA3
       BETA0=BETA3
      K=1
C     IF(NM) 9669,9668,9668
       IF ( IOP.NE.0 )  GO TO 9669
 9668 CONTINUE
       CALL SETSMG ( Z5,14,2. )
 9669  CONTINUE
       ML=MM/2
       NL=NM/2
       NLP1=NL+1
       MLP1=ML+1
 9999 SUMA=0.0
      SUMB=0.0
      MINR=MIN0(ML,NL)
   33 C=0.0
      AMIN=A(1)
      BMAX=B(1)
      FNLP=0.0
      DO 4 I=1,NL
      IF(FNA(I).LE.0.0)FNA(I)=1.
      FNLP=FNLP+FNA(I)
      SUMA=SUMA+A(I)*FNA(I)
      C=C+A(I)*A(I)*FNA(I)
      IF(A(I).LT.AMIN)AMIN=A(I)
    4 CONTINUE
      FMLP=0.0
      DO 5 I=1,ML
      IF(FNB(I).LE.0.0)FNB(I)=1.
      FMLP=FMLP+FNB(I)
      C=C+B(I)*B(I)*FNB(I)
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
      SUMB=SUMB+B(I)*FNB(I)
      IF(B(I).GT.BMAX)BMAX=B(I)
    5 CONTINUE
   54 FORMAT ( 1H0, F4.0 )
      FNLML=FNLP+FMLP
      IF(AMIN.GE.BMAX)GO TO 222
      CC=(SUMA+SUMB)/FNLML
      SUMA=SUMA/FNLP
      SUMB=SUMB/FMLP
      IF(SUMA.GT.SUMB)GO TO 9998
      GO TO 223
  222 PRINT 224
  224 FORMAT(49H0MINIMUM A IS GREATER THAN OR EQUAL TO MAXIMUM B.)
      GO TO 748
  223 PRINT 225
  225 FORMAT(41H0AVERAGE A IS NOT GREATER THAN AVERAGE B.)
      GO TO 748
C9998 IF(INPUT.NE.0)GO TO 909
 9998  IF ( BETA3.GT.0. )  GO TO 909
   55 SIGMA0=C/FNLML- CC*CC
      SIGMA0=SQRT(SIGMA0)
      MU0=(SUMA+SUMB)/2.0
      ALPHA0=MU0/SIGMA0
      BETA0=1.0/SIGMA0
      GO TO 910
  909 SIGMA0=1.0/BETA0
      MU0=ALPHA0*SIGMA0
  910 PRINT 231,IDENT
C 231 FORMAT ( 1H2,10A8 )
  231 FORMAT ( 1H2,8A10 )
      ALPH=ALPHA0
      BET=BETA0
      DO 77 K=1,LIMIT
      QQ=1.0
  301 CONST=1.
      CALL LCOM(SUM1,SUM2,NL,A,FNA,CONST,TONEX,TTWOX,THREEX,TFOURX,
     1      TFIVEX,A(NLP1),FNA(NLP1) )
      CONST=-1.
      CALL LCOM(SUM3,SUM4,ML,B,FNB,CONST,TONEY,TTWOY,THREEY,TFOURY,
     1      TFIVEY,B(MLP1),FNB(MLP1) )
      FLB=SUM1-SUM3
      FLAB=TFIVEY-TFOURY+TFIVEX+TFOURX
      FLBB=SUM4-SUM2
      FLA=TONEY-TONEX
      FLAA=TTWOY-THREEY-TTWOX-THREEX
      DELTA0=FLAA*FLBB-FLAB*FLAB
      G(K)=(FLB*FLAB-FLA*FLBB)/DELTA0
      F(K)=(FLA*FLAB-FLB*FLAA)/DELTA0
      BETA0=BETA0+F(K)
      ALPHA0=ALPHA0+G(K)
      SUM3=1.0/BETA0
      SUM4=ALPHA0/BETA0
      QQSAV(K)=QQ
   65 CONTINUE
      IF(ABS(G(K)).GE.ABS(EP1*ALPHA0).OR.ABS(F(K)).GE.ABS(EP2*BETA0))GO
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
     1TO 77
      GO TO 821
   77 CONTINUE
      PRINT 300,LIMIT
  300 FORMAT(61HOTHE DESIRED IMPROVEMENT IN ALPHA AND BETA WAS NOT MADE
     1AFTER,I4,12H ITERATIONS.)
      K=LIMIT
  821 R=0.0
      S=0.0
      T=0.0
      DO 500 J=1,NL
       L=NL+J
       IF ( FNA(L).EQ.0. )  GO TO 500
       IF ( A(L) .EQ.0.0 )  GO TO  500
       C=FNA(J)*FNA(L)/(1./A(L)-1./FNA(L) )
      U=A(J)*B1-A1
      V=U*C
      R=R+C
      S=S+V
      T=T+U*V
  500 CONTINUE
      DO 501 J=1,ML
       L=ML+J
      IF(FNB(L).EQ.0.0)GO TO 501
       IF ( B(L) .EQ.0.0 )  GO TO 501
       C=FNB(J)*FNB(L)/(1./B(L)-1./FNB(L) )
      U=B(J)*B1-A1
      V=U*C
      R=R+C
      S=S+V
      T=T+U*V
  501 CONTINUE
      C=B1*B1
      R=C*R
      S=C*S
      T=C*T
      U=AUU*ASS-AUS*AUS
      RDL= 1./U
      AUUU= ASS * RDL
      AUSS= AUU * RDL
      AUUS= -AUS * RDL
2201  CONTINUE
      PRINT 69
   69 FORMAT(6H0 A(I),17X,4HB(J))
      KK=MINR
      PRINT 722,A(1),B(1),SUM4,SUM3
  722 FORMAT(4X,3H1) ,E11.5,7X,3H1) ,E11.5,7X,3HMU=E20.14,3X,6HSIGMA=E20
     1.14 )
      DO 723 I=2,MINR
      IF(I.EQ.3)GO TO 724
      IF(I.EQ.4)GO TO 725
      PRINT 72, I, A(I),I,B(I)
   72 FORMAT(1H ,I4,2H) ,E11.5,4X,I4,2H) ,E11.5)
      GO TO 723
  724 PRINT 726,  A(I)  ,B(I),AUUU,AUUS
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
  726 FORMAT(4X,3H3) ,E11.5,7X,3H3) ,E11.5,7X,17HCOVARIANCE MATRIX,2E22.
     114)
      GO TO 723
  725 PRINT 727,A(I),B(I),AUUS,AUSS
  727 FORMAT(4X,3H4) ,E11.5,7X,3H4) ,E11.5,24X,2E22.14)
  723 CONTINUE
      IF(NL-ML)4002,4000,4004
 4004 MA=NL-ML
      DO 4001 I=1,MA
      KK=MINR+I
      IF(KK.EQ.3)GO TO 728
      IF(KK.EQ.4)GO TO 729
      PRINT 70,KK,A(KK)
   70 FORMAT(1H ,I4,2H) ,E11.5)
      GO TO 4001
  728 PRINT 730,A(KK),AUUU,AUUS
  730 FORMAT(4X,3H3) ,E11.5,28X,17HCOVARIANCE MATRIX,2E22.14)
      GO TO 4001
  729 PRINT 731,A(KK),AUUS,AUSS
  731 FORMAT(4X,3H4) ,E11.5,45X,2E22.14)
 4001 CONTINUE
      GO TO 4000
 4002 MB=ML-NL
      DO 4003 I=1,MB
      KK=MINR+I
      IF(KK.EQ.3)GO TO 732
      IF(KK.EQ.4)GO TO 733
      PRINT 71,KK,B(KK)
   71 FORMAT(22X,I4,2H) ,E11.5)
      GO TO 4003
  732 PRINT 734,B(KK),AUUU,AUUS
  734 FORMAT(25X,3H3) ,E11.5,7X,17HCOVARIANCE MATRIX,2E22.14)
      GO TO 4003
  733 PRINT 735,B(KK),AUUS,AUSS
  735 FORMAT(25X,3H4) ,E11.5,24X,2E22.14)
 4003 CONTINUE
 4000 IF(KK.GE.3)GO TO 750
      PRINT 736,AUUU,AUUS
  736 FORMAT(46X,17HCOVARIANCE MATRIX,2E22.14)
  750 IF(KK.EQ.3)PRINT 737,AUUS,AUSS
  737 FORMAT(63X,2E22.14)
      PRINT 800
  800 FORMAT(19H0NUMBER OF A VALUES,9X,8HB VALUES)
       DO 809  I=1,MINR
       I4=FNA(I)
       I6=FNB(I)
       PRINT 801,I,I4,I,I6
  809 CONTINUE
  801 FORMAT(1H ,I4,2H) ,I5  ,10X,I4,2H)   ,I5   )
      IF(NL-ML)802,149,805
  802 DO 803 I=1,MB
      KK=MINR+I
       I4=FNB(KK)
  803 PRINT 804,KK, I4
  804 FORMAT(1H ,21X,I4,2H) ,  I5 )
      GO TO 149
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
  805 DO 806 I=1,MA
      KK=MINR+I
       I4= FNA(KK)
  806 PRINT 807,KK, I4
  807 FORMAT(1H ,I4,2H) , I5 )
  149 CONTINUE
       I4=FNLP
       I6=FMLP
      PRINT 808,I4,I6
  808 FORMAT(1H0,5HTOTAL  , I6,15X,I6 )
CP    IF (INPUT.EQ.0) GO TO 99
       IF ( BETA3.LE.0. )  GO TO 99
      PRINT 151,ALPH,BET,MU0,SIGMA0
  151 FORMAT (15H0ALPHA0(INPUT)=,E21.14,15H  BETA0(INPUT)=,E21.14,6H  MU
     10=,E21.14,9H  SIGMA0=,E21.14)
       GO TO 100
   99 PRINT 150,ALPH,BET,MU0,SIGMA0
  150 FORMAT(8H0ALPHA0=E21.14,8H  BETA0=,E21.14,8H    MU0=,E21.14,9H  SI
     1GMA0=,E21.14)
  100 PRINT 98
   98 FORMAT(5H0STEP,10X,11HDELTA ALPHA,11X,10HDELTA BETA,14X,1HL)
      DO 749 I=1,K
  749 PRINT 97,I,G(I),F(I),QQSAV(I)
   97 FORMAT(1H ,I3,2X,3E21.14)
      PRINT 335,QQ,DELTAD,ALPHA0,BETA0
  335 FORMAT(9H0MAXIMUM=E20.14,8H  DELTA=,E20.14,8H  ALPHA=,E20.14,7H  B
     1ETA=,E20.14)
       DATA ( DLAB(I),I=1,8 )  /10HDISTANCE B ,10HETWEEN MU ,10H TICK MA
     1RK  ,10HS          ,10HETWEEN SIG ,10H(E8.1) )   ,10H
     2   ,10H(A10)        /
      DATA D1(1)/1.39/,D1(2)/5.99/,DRST/500./,IRSS/ 501/,IRST/1001/,
     1         PLT(1)/2H33/,PLT(2)/2H54/,ARA/3HSIG/,ARB/2HMU/,
     2         FAC/.95/,PT/1H./,ST/1H*/ ,ALM/1.000000000001/  ,
     3         XIB/625./,ETB/575./,EM/5.0/
       DATA ( T652(I),I=1,10)  /10H          ,10H          ,10H
     1           ,10H          ,10H       MU= ,10H           ,10H
     2     , 10H SIGMA=      ,10H            ,10H           /
       DATA ( T653(I),I=1,11) /10H           ,10H           ,10H
     1    ,10H          ,10H       COV ,10HARIANCE MA  ,10HTRIX
     2  ,10H          ,10H          ,10H          ,10H          /
       DATA ( T655(I),I=1,11)  / 10H          ,10H          ,10H
     1      ,10H          ,10H          ,10H          ,10H
     2  ,10H          ,10H          ,10H          ,10H          /
       DATA ( T657(I),I=1,14 ) /10H          ,10H          ,10H
     1            ,10H          ,10 H       ALP  ,10HPHA*=       ,10H
     2       , 10H    BETA*=   ,10H          ,10H    MU*=
     3    ,10H          ,10H   SIGMA*= ,10H          ,10H          /
       DATA ( T659(I),I=1,11)  /10H          ,10H          ,10H
     1   ,10H          ,10H       STE,10HP          ,10H       DELT
     2 ,10HA ALPHA    ,10H       DELT  ,10HA BETA     ,10H          L /
       DATA ( T660(I),I=1,12 )  /10H          ,10H          ,10H
     1     ,10H          ,10H          ,10H          ,10H          ,10H
     2H          ,10H          ,10H          ,10H          ,10H
     3          /
       DATA ( T661(I),I=1,3 ) /10HNO.     A ,10H       NO. ,10H      B
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
     1   /
         DATA ( T658(I),I=1,4 ) / 10H            ,10H            ,10H
     1      ,10H                /
         DATA ( T606(I),I=1,3 ) /10HTOO MANY A ,10H AND B TO  ,10H PRINT
     1    /
         DATA ( T662(I),I=1,4 )  /10HORIGIN MU= ,10H            ,10H
     1       ,10H               /
         DATA ( T663(I),I=1,4 )  /10H       SIG= ,10H
     1    ,10H            ,10H            /
         DATA ( T599(I),I=1,10 ) /10H           ,10H
     1  ,10H            ,10H            ,10H            ,10H
     2  ,10H            ,10H            ,10H            ,10H          /
         DATA ( T605(I),I=1,4 ) /10H            ,10H            ,10H
     1    ,10H            /
         DATA ( T604(I),I=1,2 ) /10H            ,10H            /
C      IF(NM) 748,608,608
       IF ( IOP.NE.0 )  GO TO 748
   608 RC= 1./ ASS
       AD1= ASS  * D1(1)                                                  EPP
       AD2= ASS  * D1(2)                                                  EPP
       SAD1= SQRT(AD1 * RDL)                                              EPP
       SAD2= SQRT(AD2 * RDL)                                              EPP
       UAD1= SQRT(AUU  * D1(1) * RDL )                                    EPP
       UAD2= SQRT(AUU  * D1(2) * RDL )                                    EPP
       XMX1= SUM4 + SAD1                                                  EPP
       XMX2= SUM4 + SAD2                                                  EPP
       XMN1= SUM4 - SAD1                                                  EPP
       XMN2= SUM4 - SAD2                                                  EPP
       SMX1= SUM3 + UAD1                                                  EPP
       SMX2= SUM3 + UAD2                                                  EPP
       SMN1= SUM3 - UAD1                                                  EPP
       SMN2= SUM3 - UAD2                                                  EPP
       DSM = 2.0 * AMAX1 (SAD1,SAD2)
       DSS = 2.0 * AMAX1 (UAD1,UAD2)
       TMP = ALOG10 (DSM)
       TNP = AINT (TMP)
       IF (TMP .LT. 0.0) TNP = TNP - 1.0
       A1 = 10. **(TMP - TNP - 1.0)
       IF(A1 .LT. 0.1) A1 = 10.0 * A1
       C1 = TNP + 1.0
       IF(A1 .LT. 0.1) C1 = C1 - 1.0
       TMP = ALOG10 (DSS)
       TNP = AINT (TMP)
       IF (TMP .LT. 0.0) TNP = TNP - 1.0
       A2 = 10. **(TMP - TNP - 1.0)
       IF(A2 .LT. 0.1) A2 = 10.0 * A2
       C2 = TNP + 1.0
       IF(A2 .LT. 0.1) C2 = C2 - 1.0
       CON = 10.** C1
       DX = .02 * CON
       IF (A1 .LT. 0.2) DX = .01 * CON
       IF (A1 .GE. 0.5) DX = .05 * CON
       CON = 10.** C2
       DY = .02* CON
       IF (A2 .LT. 0.2) DY = .01 * CON
       IF (A2 .GE. 0.5) DY = .05 * CON
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
      DX5= 5. * DX/ ALM
      DY5= 5. * DY / ALM
      TMP = AINT(AMIN1(XMN1,XMN2)/DX)
      IF (TMP .LT. 0.0) TMP = TMP - 1.0
      XMN = TMP * DX
      TMP = AINT(AMIN1(SMN1,SMN2)/DY)
      IF (TMP .LT. 0.0) TMP = TMP - 1.0
      SMN = TMP * DY
      XMX = AMAX1(XMX1,XMX2)
      TMP = XMX/DX
      TMPI = AINT(TMP)
      IF (TMPI .LT. 0.0) TMPI = TMPI - 1.0
      IF(TMPI .NE. TMP) XMX = (TMPI + 1.0) * DX
      SMX = AMAX1(SMX1,SMX2)
      TMP = SMX/DY
      TMPI = AINT(TMP)
      IF (TMPI .LT. 0.0) TMPI = TMPI - 1.0
      IF(TMPI .NE. TMP) SMX = (TMPI + 1.0) * DY
      DGX= XMX - XMN
      DGY= SMX - SMN
      CI1 = 1024. / 8.94
      CI2 = 1024. / 7.42
      CI3 = CI2 / CI1
      R1 = DSM / (EM * CI1)
      R2 = DSS / (EM * CI2)
      RI1 = 1.0 / R1
      RI2 = 1.0 / R2
      XICON= XIB - RI1 * SUM4
      ETCON= ETB + RI2 * SUM3
      ITA = 1
      IF(DSM .LT. 1.0) ITA = ABS((C1-2.0)*ALM)
      ITO = 1
      IF(DSS .LT. 1.0) ITO = ABS((C2-2.0)*ALM)
C     XI VALUE FOR TICK MARKS
      NTICX = ALM * DGX/DX + 2.
      DO 405 I = 1, NTICX
      Z1= I - 1
      XLAB1(I) = (XMN + Z1* DX)
      IF(I.GT.5) GO TO 403
      IF(ABS(AMOD(XLAB1(I),DX5)).LT..5*DX) J1=I
  403 ITX(I) = XICON + RI1 * XLAB1(I)
  405 CONTINUE
C     A FORMAT FOR VALUE OF X AXIS
      IF(J1.EQ.1) J1 = 6
      J= J1
      NTX =(NTICX -J1)/5+1
      DO 407 I=1,NTX
       I5=2
       IXL(I)=7
       IF ( ITA.GT.4 )  ITA=4
       CALL FMTSG (Z5,I5,IXL(I),ITA,XLAB1(J),XLAB2(I) )
      J= J + 5
  407 CONTINUE
C     ETA VALUE FOR TICK MARKS
      NTICY = ALM * DGY / DY + 2.
      DO 410 I=1, NTICY
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
      Z1= I - 1
      YLAB1(I) = (SMN + Z1* DY)
      IF(I.GT.5) GO TO 408
      IF(ABS(AMOD(YLAB1(I),DY5)).LT..5*DY) J2=I
  408 ITY(I) = ETCON - RI2 * YLAB1(I)
  410 CONTINUE
C     A FORMAT FOR VALUE OF Y AXIS
      IF(J2.EQ.1) J2 = 6
      J= J2
      NTY =(NTICY -J2)/5+1
      DO 412 I=1,NTY
      I5=2
      IYL(I)=7
      IF ( ITO.GT.4 )  ITO=4
      CALL FMTSG ( Z5,I5,IYL(I),ITO,YLAB1(J),YLAB2(I) )
      J= J + 5
  412 CONTINUE
      AK1 = EM**2 * U / (4.0 * AUU * ASS)
      AK2 = AK1 * D1(1) / D1(2)
      AC = AUU * ASS
      BSAC = CI3*AUS / SQRT (AC)
      UAC2 = U /(AC * CI1 * CI1)
C     XI + ETA FOR  .95
      XIMIN = XICON + RI1 * XMN2
      XIN= RI1 * (XMX2-XMN2)/DRST
      XINOW = XIMIN
      DO 420 I=1,IRST, 2
      XIDIF = XINOW - XIB
      IX(I) = XINOW
      IX(I+1) = XINOW
      PARTL = ETB + BSAC * XIDIF
      RDC= AK1 - UAC2 * XIDIF **2
      IF(RDC.GE.0.) GO TO 418
      PARTR= 0.
      GO TO 419
  418 PARTR = CI2 * SQRT (RDC)
  419 IY(I) = PARTL + PARTR
      IY(I+1) = PARTL - PARTR
      XINOW = XINOW + XIN
  420 CONTINUE
C     XI + ETA FOR  .50
  421 XJMIN = XICON + RI1 * XMN1
      XJN= 2. * RI1 * (XMX1-XMN1)/DRST
      XJNOW = XJMIN
      DO 425 I=1,IRSS,2
      XJDIF = XJNOW - XIB
      JX(I) = XJNOW
      JX(I+1) = XJNOW
      PARTL = ETB + BSAC * XJDIF
      RDD= AK2 - UAC2 * XJDIF **2
      IF(RDD.GE.0.) GO TO 423
      PARTR= 0.
      GO TO 424
  423 PARTR = CI2 * SQRT (RDD)
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
424 JY(I) = PARTL + PARTR
    JY(I+1) = PARTL - PARTR
    XJNOW = XJNOW + XJN
425 CONTINUE
426 IRSV = IRSS + 1
    JX(IRSV) = XIB
    JY(IRSV) = ETB
     CALL PAGEG ( Z5,0,1,1 )
     I7=0
     I9=3071
     CALL LEGNDG ( Z5,I7,I9,1,2H    )
     XPI=31.
     YPI=54.
     CALL SETSMG(Z5,14,1. )
     XP=XPI
     YP=3071.-1.5*YPI
     CALL FMTSG ( Z5,3,12,6,SUM4,T652(6) )
     CALL FMTSG ( Z5,3,12,6,SUM3,T652(9 ) )
     CALL LEGNDG ( Z5,XP,YP,92,T652(1) )
     YP=YP-2.*YPI
     CALL FMTSG ( Z5,3,18,6,AUUU,T653(8) )
     CALL FMTSG ( Z5,3,18,6,AUUS,T653(10) )
     CALL LEGNDG ( Z5,XP,YP,110,T653(1) )
     YP=YP-2.*YPI
      CALL FMTSG ( Z5,3,18,6, AUUS,T655(8) )
      CALL FMTSG ( Z5,3,18,6, AUSS,T655(10) )
     CALL LEGNDG ( Z5,XP,YP,110,T655(1) )
     YP=YP-2.*YPI
 86 FORMAT ( A2,R8 )
     CALL FMTSG ( Z5,3,12,6, ALPH  ,TEMP(1) )
     T657(  7)= TEMP(1)
     ENCODE ( 10,86,T657( 8) ) TEMP(2),T657( 8 )
     CALL FMTSG ( Z5,3,12,6, BET   ,TEMP(1) )
     T657(  9)= TEMP(1)
     ENCODE ( 10,86,T657(10) ) TEMP(2),T657( 10)
     CALL FMTSG ( Z5,3,12,6, MU0   ,TEMP(1) )
     T657( 11)= TEMP(1)
     ENCODE ( 10,86,T657(12) ) TEMP(2),T657( 12)
     CALL FMTSG ( Z5,3,12.6, SIGMA0,TEMP(1) )
     T657( 13)= TEMP(1)
     ENCODE ( 10,86,T657(14) ) TEMP(2),T657( 14)
     XP=.5*XPI
     CALL LEGNDG ( Z5,XP,YP,132,T657(1) )
     XP=XPI
     YP=YP-2.*YPI
    IF(NL.LT.3.AND.ML.LT.3) GO TO 650                                  EPP
    IF(NL.GT.50.OR.ML.GT.50) GO TO 645
    MNN= MIND(NL,ML)                                                   EPP
     CALL LEGNDG ( Z5,XP,YP,110,T659(1) )
     YP=YP-2.*YPI
     DO 634  I=1,K
     CALL FMTSG ( Z5,1,10,0,I,T660(5) )
     CALL FMTSG ( Z5,3,18,6,G(I),T660(7) )
     CALL FMTSG ( Z5,3,18,6,F(I),T660(9) )
     CALL FMTSG ( Z5,3,18,6,QQSAV(I),T660(11) )
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
      CALL LEGNDG ( Z5,XP,YP,120,T660(1) )
      YP=YP-2.*YPI
  634 CONTINUE
  427 CONTINUE
      CALL SETSMG( Z5,14,2. )
      I7=0
      I9=3071
      CALL LEGNDG ( Z5,I7,I9,1,1H    )
      CALL SETSMG ( Z5,14,1. )
      YP=I9
      YP=YP-1.5*YPI
      CALL LEGNDG ( Z5,XP,YP, 80,IDENT(1) )
      YP=YP-2.*YPI
      CALL LEGNDG ( Z5,XP,YP,30,T661(1) )
      YP=YP-2.*YPI
      DO 6350 I=1,MNN
     I1=FNA(I)
      CALL FMTSG ( Z5,1,3,0,I1     ,T658(1) )
      CALL FMTSG ( Z5,3,12,5,A(I),TEMP(1) )
      ENCODE ( 10,80,T658(1) ) T658(1),TEMP(1)
   80 FORMAT ( A3,A7)
      ENCODE (  5,81,T658(2) ) TEMP(1),TEMP(2)
   81 FORMAT ( R3,A2 )
      I1=FNB(I)
      CALL FMTSG ( Z5,1,4,0,I1     ,TEMP(3) )
      ENCODE (  9,82,T658(2) ) T658(2),TEMP(3)
   82 FORMAT ( A5,A4 )
      CALL FMTSG ( Z5,3,12,5,B(I),T658(3 ) )
      CALL LEGNDG ( Z5,XP,YP,40,T658(1) )
      YP= YP-1.5*YPI
 6350 CONTINUE
      MNN= MNN + 1                                                    EPP
      IF(NL- ML) 635,650,640                                          EPP
  635 CONTINUE
      DO 6351  I=MNN,ML
      I1=FNB(I)
       CALL FMTSG ( Z5,1,4,0,I1     ,TEMP(1) )
      ENCODE (10,83,T605(2) ) T605(2),TEMP(1)
   83 FORMAT ( A5,A5 )
      CALL FMTSG ( Z5,3,12,5,B(I),T605(3) )
      CALL LEGNDG ( Z5,XP,YP,40,T605(1) )
      YP= YP-1.5*YPI
 6351 CONTINUE
      GO TO 650
  640 CONTINUE
      DO 6405  I=MNN,NL
      I1=FNA(I)
      CALL FMTSG ( Z5,1,3,0,I1     ,TEMP(1) )
      CALL FMTSG ( Z5,3,12,5,A(I),TEMP(3) )
      ENCODE ( 10,84,T604(1) ) TEMP(1),TEMP(3)
   84 FORMAT ( A3,A7 )
      ENCODE ( 5,85,T604(2) ) TEMP(3),TEMP(4)
   85 FORMAT ( R3,A2 )
      CALL LEGNDG ( Z5,XP,YP,20,T604(1) )
      YP= YP-1.5*YPI
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
 6405 CONTINUE
      GO TO 650                                                    EPP
  645 CONTINUE
       CALL LEGNDG ( Z5,XP,YP,30,T606(1) )
       YP=YP-YPI
  650 CONTINUE
       CALL SETSMG ( Z5,14,2. )
C      PLOT AXIS
       I7=(FLOAT(ITX(1))*4095.)/1023.
       I9=(1023.-FLOAT(ITY(1)))*3071./1023.
       I11= ( FLOAT(ITX(NTICX))*4095.)/1023.
       CALL SEGMTG ( Z5,1,I7,I9,I11,I9 )
       I13= (1023.-FLOAT(ITY(NTICY)))*3071./1023.
       CALL SEGMTG ( Z5,1,I7,I13,I7,I9 )
C     TICK MARKS ON X AXIS
       KB = ITY(1) - 3
       KC = ITY(1) + 3
        I9= ((1023.-FLOAT(KB))* 3071.)/1023.
        I13= ((1023.-FLOAT(KC))*3071.)/1023.
       DO 430 I = 1,NTICX
       KA = ITX(I)
        I7= (FLOAT(KA)*4095.)/1023.
        CALL SEGMTG ( Z5,1,I7,I9,I7,I13 )
  430 CONTINUE
C     TICK MARKS ON Y AXIS
       KA = ITX(1) - 3
       KC = ITX(1) + 3
        I7= (FLOAT(KA)*4095.)/1023.
        I11= (FLOAT(KC)*4095.)/1023.
       DO 435 I = 1,NTICY
       KB = ITY(I)
        I9= ((1023.-FLOAT(KB))*3071.)/1023.
        CALL SEGMTG ( Z5,1,I7,I9,I11,I9 )
  435 CONTINUE
      DO 437 I=1,4
      CHM(I)= DLAB(I)
  437 CHS(I)= DLAB(I)
      CHS(2)= DLAB(5)
      ICT=  10
       INP=10
       ENCODE ( ICT,DLAB(6),CHM(5) ) DX
       DECODE ( INP,DLAB(8),CHM(5)) CHM(5)
       ENCODE ( ICT,DLAB(6),CHS(5) ) DY
       DECODE ( INP,DLAB(8),CHS(5) ) CHS(5)
C      LABEL X AXIS
       NY=ITY(1)+10
       J= J1
       DO 440 I=1,NTX
       NX = ITX(J) - 4 * IXL(I) + 4
       J= J+5
        I7= ( FLOAT( NX ) *4095.)/1023. +2.*XPI
        I9=(( 1023.-FLOAT(NY))*3071.)/1023.
        CALL LEGNDG ( Z5,I7,I9,IXL(I),XLAB2(I) )
  440 CONTINUE
      NY= NY + 20
      JTX= NTICX/3
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
      I7= (FLOAT(ITX(JTX))*4095.)/1023.
      I9= ((1023.-FLOAT(NY))*3071.)/1023.
      CALL LEGNDG ( Z5,I7,I9, 50      ,CHM(1) )
C     LABEL Y AXIS
      J= J2
      DO 445 I = 1,NTY
      NX = ITX(1) - 8 * (IYL(I) + 1)
      NY = ITY(J)
      J= J+5
      I7= (FLOAT(NX)*4095.)/1023.
      I9=((1023.-FLOAT(NY))*3071.)/1023.
      CALL LEGNDG ( Z5,I7,I9,IYL(I),YLA82 (I) )
  445 CONTINUE
      I7= (FLOAT(5)*4095.)/1023.
      I9= ((1023.-FLOAT(850))*3071.)/1023.
      CALL LEGNDG ( Z5,I7,I9,40,CHS(1))
      I7= (FLOAT(85)*4095.)/1023.
      I9=((1023.-FLOAT(865))*3071.)/1023.
      CALL LEGNDG ( Z5,I7,I9,10,CHS(5) )
      CALL FMTSG ( Z5,3,12,5,XMN,T662(3) )
      CALL FMTSG ( Z5,3,12,5,SMN,T663(3) )
      CALL SETSMG (Z5,14,1. )
      YP=I9
      YP=YP-1.5*YPI
      CALL LEGNDG ( Z5,XP,YP,40,T662(1) )
      YP=YP-1.5*YPI
      CALL LEGNDG ( Z5,XP,YP,40,T663(1) )
      YP=YP-1.5*YPI
      CALL SETSMG (Z5,14,2. )
C     PLOT FOR  .95
      DO 450 I = 1, IRST
      I7= (FLOAT(IX(I))*4095.)/1023.
      I9=((1023.-FLOAT(IY(I)))*3071.)/1023.
      CALL LEGNDG ( Z5,I7,I9,1,PT )
  450 CONTINUE
C     PLOT FOR  .50
      DO 460 I = 1,IRSS
      I7= (FLOAT(JX(I))*4095.)/1023.
      I9=((1023.-FLOAT(JY(I)))*3071.)/1023.
      CALL LEGNDG ( Z5,I7,I9,1,PT )
  460 CONTINUE
      I7= (FLOAT(JX(IRSV))*4095.)/1023.
      I9=((1023.-FLOAT(JY(IRSV)))*3071.)/1023.
      CALL LEGNDG ( Z5,I7,I9,1,ST )
      NXX = ITX(1) - 10
      NYY = ITY(NTICY) -15
      I7= (FLOAT(NXX)*4095.)/1023.
      I9=((1023.-FLOAT(NYY))*3071.)/1023.
      CALL LEGNDG ( Z5,I7,I9,3,ARA )
      NXX = ITX(NTICX) +15
      NYY = ITY(1)
      I7= (FLOAT(NXX)*4095.)/1023.
      I9=((1023.-FLOAT(NYY))*3071.)/1023.
      CALL LEGNDG ( Z5,I7,I9,2,ARB )
```

(cont'd on next page)

APPENDIX 9A (cont'd)

```
  748  CONTINUE
1001  CONTINUE
       RETURN
      END
```

APPENDIX 9B
COMPUTER PROGRAM 9-2

(Einbinder's Maximum Likelihood Estimation of Weibull Parameters)

INPUT GUIDE (8/79)

FOR WEIBULL SENSITIVITY PROGRAM

CARD SET	SYMBOL	CARD COLUMNS	FORMAT	DESCRIPTION
1	IDENT	1-80	8A10	Title or identifying information.
2	N	1-3	I3	Sample Size (150 Max).
3	S(I)	1-10, 11-20, etc.	7F10.0	Stress levels, 7 per card, I = 1, N.
4	U(I)	1-80	80I1	N Responses: Positive Response = 1 Negative Response = 0.
5	EPILSON	1-10	F10.0	Convergence Accuracy desired (0.00001 is usually sufficient).
	ICOUNT	11-15	I5	Max number of iterations, Default = 25.
6	IGAM	1	I1	Option for GAMMA (See Note 1) = 1 Search from ASTART to max admissible value. = 2 Search from ASTART to LASTG. = 3 Use fixed GAMMA. Specify values in card set 9.
7	ISTART	1	I1	Quantile procedure for estimating starting values for iterative solution. = 0 Built-in quantiles are used, Viz, P1 = 0.15, XP1 = Xmin1, P2 = 0.85, XP2 = Xmax0. = 1, read quantiles on card set 11.
	IREFL	2	I1	Type of Weibull Distribution = 0, Standard Weibull = 1, Reflected Weibull.
	NCL	3	I1	Number of confidence coefficients for interval estimates of reliability (one-sided) and/or quantiles (two-sided), up to 5.
	NCR	4-5	I2	Number of reliability boundary values (up to 20).
	NPL	6-7	I2	Number of quantiles (percentage points) of response function (up to 30).
	NGAM	8-9	I2	Number of gamma values when IGAM = 3.

(Omit card set 8 if IGAM = 3)

(cont'd on next page)

APPENDIX 9B (cont'd)

CARD SET	SYMBOL	CARD COLUMNS	FORMAT	DESCRIPTION
8	ASTART	1-10	F10.0	Minimum value of gamma search interval for standard Weibull; maximum value for reflected Weibull.
	ASTEP	11-20	F10.0	GAMMA step size for search option.
	LASTG	21-30	F10.0	Maximum value of GAMMA search interval for standard Weibull, minimum value for reflected; NOTE: not required if IGAM = 3.
9	GAMMA (Required if IGAM = 3)	1-70	7F10.0	Values of Gamma.
10	COEF(I)	1-10, 11-20, etc.	5F10.0	Confidence coefficients, I = 1, NCL.
(Omit Card Set 10 if NCL = 0)				
11	CR(I)	1-10 11-20	7F10.0	Reliability boundary values, 7 per card, I = 1, NCR.
(Omit Card Set 11 if NCR = 0)				
12	PL(I)	1-10	7F10.0	Response function probability levels corresponding to desired quantiles, Lp, 7 per card, I = 1, NPL.
(Omit Card Set 12 if NPL = 0)				
13	P1	1-10	F10.0	Lower response probability for estimating starting values of parameters.
	XP1	11-20	F10.0	Quantile (percentage point) corresponding to P1.
	P2	21-30	F10.0	Upper response probability.
	XP2	31-40	F10.0	Corresponding quantile.

NOTE 1: A three-parameter covariance matrix is computed if gamma is estimated by searching for max likelihood using option IGAM = 1 or 2. A two-parameter (theta, alpha) covariance matrix is computed if gamma is specified as known (IGAM = 3).

(cont'd on next page)

APPENDIX 9B (cont'd)

OUTPUT

Sample Problem AORS 17

I	STIMULUS	RESPONSE
1	32.0000	1
2	32.0000	1
3	32.0000	1
4	32.0000	0
5	48.0000	1
6	48.0000	1
7	48.0000	1
8	48.0000	1
9	40.0000	1
10	40.0000	1
11	40.0000	1
12	40.0000	1
13	20.0000	0
14	30.0000	1
15	30.0000	1
16	30.0000	0
17	39.0000	1
18	39.0000	1
19	39.0000	1
20	39.0000	1
21	34.5000	1
22	34.5000	1
23	34.5000	1
24	34.5000	1
25	27.2500	0
26	30.8800	1
27	30.8800	1
28	30.8800	1
29	30.8800	1
30	29.0600	1
31	29.0600	0

.0000100025000

3

001 3 6 1

25.0000

.9500

30.0000 34.0000 38.0000

.1000 .5000 .8400 .9000 .9500 .9900

XMIN1 = 29.0600 XMAX0 = 32.0000

(cont'd on next page)

APPENDIX 9B (cont'd)

WEIBULL QUANTAL RESPONSE ESTIMATION

P(X.LE.XMIN1)	P(X.GT.XMAX0)	ALPHA START	LAMBDA START	GAMMA	ITERATIONS	LAST DEL ALPHA	LAST DEL LAMBDA
.503413	.096093	.451106E+01	-.813778E+01	.250000E+02	10	-.465286E-04	.260151E-05

	THETA	ALPHA	GAMMA	LOG L
1	.47685E+01	.22174E+01	.25000E+02	-.434236E+01

STANDARD WEIBULL DISTRIBUTION

MAX LIKELIHOOD ESTIMATES ARE

LAMBDA= .313154E-01 ALPHA= .221739E+01
GAMMA= .250000E+02 THETA= .476853E+01
MAX LOG L= -.434236E+01
L50= 29.0420
MEAN= 29.2233 STANDARD DEVIATION= 2.0122

FISHER INFORMATION MATRIX FOR THETA, ALPHA, GAMMA

1.89084E+00 -8.55880E-01
-8.55880E-01 1.00220E+00

NO CONVERGENCE IN IMPRUV. MATRIX IS NEARLY SINGULAR

ASYMPTOTIC COVARIANCE MATRIX FOR THETA, ALPHA, GAMMA

8.62133E-01 7.36266E-01
7.36266E-01 1.62658E+00

PRODUCT OF INFO AND COVARIANCE MATRICES

1.00000E+00 0.
-3.55271E-15 1.00000E+00

GENERALIZED VARIANCE= .86024E+00

ASYMPTOTIC RELIABILITY ESTIMATES

C	REL	VAR R	SIG R	C COEF	LCL
30.0000	.329287E+00	.210884E-01	.145218E+00	.950	.904221E-01
34.0000	.167455E-01	.191229E-02	.437298E-01	.950	0.
38.0000	.968084E-04	.909321E-06	.953583E-03	.950	0.

(cont'd on next page)

WEIBULL QUANTILE ESTIMATES

P	L(P)	SIG LP	C COEF	LCL	UCL
.1000	.26728E+02	.12463E+01	.950	.24286E+02	.29171E+02
.5000	.29042E+02	.10692E+01	.950	.26946E+02	.31138E+02
.8400	.31266E+02	.98202E+00	.950	.29342E+02	.33191E+02
.9000	.31946E+02	.12491E+01	.950	.29498E+02	.34394E+02
.9500	.32821E+02	.17489E+01	.950	.29393E+02	.36249E+02
.9900	.34495E+02	.29864E+01	.950	.28642E+02	.40348E+02

(cont'd on next page)

APPENDIX 9B (cont'd)

```
*DECK,SEMART
      PROGRAM SEMART(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
C
C     ********* INPUT GLOSSARY ****************
C
C
C     SYMBOL                        DESCRIPTION
C    **********    ********************************
C
C      IDENT        TITLE
C      N            SAMPLE SIZE
C      S(I)          STRESSES
C     U(I)          QUANTAL RESPONSES, 1=POSITIVE RESPONSE
C                                      0= NO RESPONSE
C      EPILSON      CONVERGENCE FACTOR, NORMALLY = .00001
C      ICOUNT       MAX NUMBER OF ITERATIONS, DEFAULT = 25
C      IGAM         LOCATION PARAMETER OPTION,
C                     1= ESTIMATE 3 PARAMETERS, DOMAIN OF GAMMA IS FROM
C                        ASTART TO XMIN1
C                     2= ESTIMATE 3 PARAMETERS, DOMAIN OF GAMMA IS ASTART
C                        TO LASTG
C                     3= ESTIMATE 2 PARAMETERS ASSUMING GAMMA IS KNOWN
C      ISTART       OPTION FOR MATCHING PERCENTAGE POINTS,
C                     0= DEFAULT OPTION, P1=.15, XP1=XMIN1
C                                        P2=.85,XP2=XMAX0
C                     1= INPUT P1,XP1,P2,XP2.
C      IREFL        0= FIT STANDARD WEIBULL DISTRIBUTION
C                   1= FIT REFLECTED WEIBULL DISTRIBUTION
C      NCL          NUMBER OF CONFIDENCE COEFFICIENTS
C     NCR           NUMBER OF RELIABILITY BOUNDARIES DESIRED
C      NPL          NUMBER OF PERCENTAGE POINTS DESIRED
C      NGAM         NUMBER OF GAMMA VALUES  ASSUMED FOR 2 PARAMETER
C                     ESTIMATION (IGAM=3)
C      ASTART       MIN GAMMA FOR LOCATION PARAMETER INTERVAL
C      ASTEP        GAMMA STEP SIZE FOR  3 PARAMETER ESTIMATION
C      LASTG        MAX GAMMA OF SEARCH DOMAIN, DEFAULT = XMIN1 FO IGAM=1
C      COEF(I)      CONFIDENCE COEFFICIENTS
C      CR(I)        RELIABILITY BOUNDARY VALUES
C      PL(I)        PERCENTILES
C      P1,XP1       100*P1 PERCENTAGE ASSOCIATED WITH XP1
C      P2,XP2       100*P2 PERCENTAGE ASSOCIATED WITH XP2
      INTEGER U
      REAL LAMBDA
      REAL LAMBDA2
      REAL LASTG
      DIMENSION ST(150),CR(20),COEF(5),PL(30),SS(150)
      DIMENSION S(150),U(150),FB(150),H(150),PHA(150),PHL(150),
     1A(150), B(150), C(150), D(150), E(150), F(150)
      DIMENSION PARAM(150,4),BF(3,6),COV(3,3),V(150),SLN(150),P(150)
      DIMENSION IDENT(8),GAMMA(14)
      COMMON/BL1/V,SLN,P,S
      COMMON/BL2/NCR,NCL,CR,COEF,IREFL,ASTART
    1 FORMAT ( I3)
    2 FORMAT ( /F10.4)
    3 FORMAT (80I1)
    4 FORMAT(F10.6,I5)
    6 FORMAT(1H1,T30,*WEIBULL QUANTAL RESPONSE ESTIMATION*,/
     1T100,*LAST*,T115,*LAST*,/1X,*P(X.LE.XMIN1)*,3X,*P(X.GT.XMAX0)*,
     15X,*ALPHA START*,5X,*LAMBDA START*,4X,*GAMMA*,7X,*ITERATIONS*,
     1T100,*DEL ALPHA*,T115,*DEL LAMBDA*)
  500 CONTINUE
      READ(5,11) IDENT
```

(cont'd on next page)

APPENDIX 9B (cont'd)

```
   11 FORMAT(8A10)
      IF(EOF(5).NE.0) GO TO 111
      WRITE(6,12) IDENT
   12 FORMAT(1H1,8A10/)
      READ ( 5,1) N
      READ (5,2) (S(I),I=1,N)
      READ (5,3) (U(I),I=1,N)
      WRITE(6,1004)
 1004 FORMAT(1H0,T6,*I*,T15,*STIMULUS*,T30,*RESPONSE*/)
      DO 20 I = 1,N
      WRITE(6,1002) I,S(I),U(I)
 1002 FORMAT(3X,I3,4X,F10.4,13X,I1)
      U(I)=-FLOAT(U(I)) +1.0001
   20 CONTINUE
      READ (5,4) EPILSON,ICOUNT
      IF(ICOUNT.EQ.0) ICOUNT=25
      WRITE (6,1003) EPILSON, ICOUNT
 1003 FORMAT ( 1H , F10.8, I5)
      READ (5,1111) IGAM
      WRITE (6,1044) IGAM
      READ (5,1111) ISTART, IREFL, NCL, NCR, NPL, NGAM
      IF (IGAM.NE.3) NGAM=1
1111  FORMAT(3I1,3I2,10I1)
      WRITE(6,1044) ISTART,IREFL,NCL,NCR,NPL,NGAM
1044  FORMAT(1H ,3I1,3I2,10I1)
      IF(IGAM.EQ.3) GO TO 667
      READ(5,2) ASTART,ASTEP,LASTG
      WRITE(6,63) ASTART,ASTEP,LASTG
      GO TO 767
   63 FORMAT(1H ,7(F10.4,2X))
667   READ (5,2) (GAMMA(I), I=1, NGAM)
      WRITE(6,63)(GAMMA(I), I=1,NGAM)
767   CONTINUE
      IF(NCL.EQ.0) GO TO 59
      READ(5,2) (COEF(I),I=1,NCL)
      WRITE(6,63) (COEF(I),I=1,NCL)
   59 IF(NCR.EQ.0) GO TO 61
      READ(5,2)(CR(I),I=1,NCR)
      WRITE(6,63)(CR(I),I=1,NCR)
   61 IF(NPL.EQ.0) GO TO 62
      READ(5,2)(PL(I),I=1,NPL)
      WRITE(6,63)(PL(I),I=1,NPL)
   62 CONTINUE
      IF(IREFL.EQ.0) GO TO 15
      IF (IGAM.NE.3) GO TO 64
      ASTART= GAMMA(1)
      DO 640 I=1, NGAM
      GAMMA(I)=2.*ASTART-GAMMA(I)
640   CONTINUE
      WRITE(6,66) ASTART
66    FORMAT(1H ,*REFLECTION COORDINATE=*,F10.4)
64    CONTINUE
      CALL RFLECT (S,U,N,ASTART,IGAM,ASTART,LASTG)
   15 CONTINUE
      IF (ISTART.EQ.0) GO TO 65
      READ(5,2) P1,XP1,P2,XP2
      WRITE (6,1045) P1,XP1,P2,XP2
65    CONTINUE
      ASTOP = 999999
      DO 709 I = 1,N
      IF ( U(I) .EQ. 1 ) GO TO 709
      IF ( ASTOP - S(I))  709,709,707
  707 ASTOP = S(I)
  709 CONTINUE
      ALAST = 0
```

(cont'd on next page)

APPENDIX 9B (cont'd)

```
      DO 777 I = 1,N
      IF ( U(I) .EQ. 0 ) GO TO 777
      IF( S(I) - ALAST ) 777,777,776
  776 ALAST = S(I)
  777 CONTINUE
      WRITE(6,1043)ASTOP,ALAST
1043  FORMAT(1H ,*XMIN1=*,F10.4,5X,*XMAXO=*,F10.4)
      IF(IGAM.NE.3) WRITE(6,6)
      DO 10 I=1,N
      ST(I)=S(I)
   10 CONTINUE
      NGM=0
      GO TO (55,56,57),IGAM
   55 BSTOP=ASTOP
      GO TO 58
   56 BSTOP=LASTG
      IF (BSTOP.GT. ASTOP) BSTOP=ASTOP
      GO TO 58
57    CONTINUE
      IF(NGM.EQ.NGAM) GO TO 109
      W=GAMMA(NGM+1)+.0000001
      BSTOP=ASTOP
   58 CONTINUE
      IF(NGM.NE.0) GO TO 499
      IF (ISTART.EQ.0) GO TO 54
      IF(IREFL.EQ.0) GO TO 53
      X1=2.*ASTART-XP2
      X2=2.*ASTART-XP1
      XP1=X1
      XP2=X2
      PP1=1.-P1
      PP2=1.-P2
      P1=PP2
      P2=PP1
      GO TO 53
   54 CONTINUE
      P1=.15
      XP1=ASTOP
      P2=.85
      XP2=ALAST
   53 CONTINUE
 1045 FORMAT (1H ,6(F10.4,3X))
      IF(ASTOP.LT.ALAST) GO TO 499
      WRITE(6,1046)
 1046 FORMAT(1H0,10X,*DEGENERATE CASE. NO ZONE OF MIXED RESULTS.
     1XMIN1.GT.XMAX0*)
      GO TO 500
  499 CONTINUE
      K=0
      IF(IGAM.EQ.3) GO TO 503
      W = ASTART-ASTEP
  501 W = W + ASTEP + .0000001
  503 CONTINUE
      IF(W.GT.BSTOP) GO TO 60
      SP1=XP1-W
      SP2=XP2-W
      K=K+1
      DO 306  I=1,N
      S(I) = ST(I) - W
      SS(I) = S(I)
  306 CONTINUE
      LA = 1
  155 CONTINUE
      ALPHA=ALOG(ALOG(1.-P1)/ALOG(1.-P2))/ALOG(SP1/SP2)
      AL=ALOG(-ALOG(1.-P1))-ALPHA*ALOG(SP1)
```

(cont'd on next page)

APPENDIX 9B (cont'd)

```
      IF ( LA .EQ. 0 ) GO TO 146
      LA = 0
      AA=AL
      BB = ALPHA
      DO 144 I = 1,N
      IF (ST(I) .LE.W) GO TO 144
      SLG=AA+BB*ALOG(S(I))
      S(I)=EXP(SLG)
  144 CONTINUE
      SP1=AA+BB*ALOG(SP1)
      SP1=EXP(SP1)
      SP2=AA+BB*ALOG(SP2)
      SP2=EXP(SP2)
      GO TO 155
  146 CONTINUE
      ALPHA2 = ALPHA*BB
      ALAM=ALPHA*AA+AL
      LAMBDA2=EXP(ALAM)
      LAMBDA=EXP(AL)
      ITER=0
   25 DO 30 I = 1,N
      IF(ST(I) .LE.W) GO TO 30
      FB(I) = 0.
      IF( S(I)**ALPHA*LAMBDA .GT. 100 ) GO TO 27
      FB(I) = EXP(-S(I)**ALPHA* LAMBDA )
   27 CONTINUE
      IF(U(I).EQ.1) GO TO 28
      H(I) = (( 1-U(I)) * FB(I))/(1-FB(I)) -  U(I)
      PHA(I) = (( U(I)-1)*LAMBDA *S(I)**ALPHA *ALOG(S(I)) * FB(I))
     1   / (1-FB(I)) **2
      PHL(I) = (( U(I) - 1) * S(I)**ALPHA * FB(I))/( 1-FB(I)) ** 2
      GO TO 591
   28 CONTINUE
      H(I)=-1
      PHA(I)=0
      PHL(I)=0
  591 CONTINUE
      B(I)=S(I)**ALPHA*H(I)
      A(I) = B(I) *ALOG(S(I))
      C(I) = S(I)**ALPHA*ALOG(S(I))*(H(I)*ALOG(S(I))+PHA(I) )
      D(I) = S(I)**ALPHA*ALOG(S(I))*PHL(I)
      E(I) = S(I) **ALPHA*(H(I) *ALOG(S(I)) + PHA(I))
      F(I) = S(I) ** ALPHA * PHL(I)
   30 CONTINUE
      AT = 0
      BT = 0
      CT = 0
      DT = 0
      ET = 0
      FT = 0
      DO 40 I = 1,N
      IF (ST(I) .LE. W) GO TO 40
      AT = AT + A(I)
      BT = BT+ B(I)
      CT = CT + C(I)
      DT = DT + D(I)
      ET = ET + E(I)
      FT = FT + F(I)
   40 CONTINUE
      DET = CT*FT-DT*ET
      DELTAH = ( -FT*AT+DT*BT)/ DET
      DELTAK = (-CT *BT + ET*AT) / DET
      ITER=ITER + 1
      IF ( ABS(DELTAH).LE.EPILSON.AND.ABS(DELTAK).LE.EPILSON) GO TO 50
      IF ( ABS(DELTAH/ALPHA) .GT. .1 ) GO TO 8111
```

(cont'd on next page)

APPENDIX 9B (cont'd)

```
      ALPHA = ALPHA + DELTAH
      GO TO 8444
 8111 ALPHA = ALPHA + .1 * ABS ( ALPHA ) * ( DELTAH/ABS(DELTAH))
 8444 CONTINUE
      IF ( ABS( DELTAK/LAMBDA) .GT. .1 ) GO TO 8333
      LAMBDA = LAMBDA + DELTAK
      GO TO 8445
 8333 LAMBDA = LAMBDA + .1*ABS(LAMBDA)*(DELTAK/ABS(DELTAK))
 8445 CONTINUE
      A2=ALPHA2
      B2= LAMBDA2
      ALPHA2 = ALPHA*BB
      ALAM=ALPHA*AA+ALOG(LAMBDA)
      LAMBDA2=EXP(ALAM)
      DELALF=A2-ALPHA2
      DELAMB=B2-LAMBDA2
      IF(ITER.EQ.ICOUNT) GO TO 50
      GO TO 25
   50 CONTINUE
      ALIKE=0.
      DO 93 I=1,N
      IF(ST(I) .LE.W) GO TO 93
      YL=ALAM+ALOG(SS(I))*ALPHA2
      IF(U(I).EQ.1.)GO TO 94
      IF(YL.GT.5.)GO TO 93
      GO TO 93
   94 ALIKE=ALIKE=ALIKE-EXP(YL)
   93 CONTINUE
      PARAM(K,1)=EXP(-ALAM/ALPHA2)
      PARAM(K,2)=ALPHA2
      PARAM(K,3)=W
      PARAM(K,4)=ALIKE
  502 DO 279 I=1,N
      S(I)=ST(I)
  279 CONTINUE
      W = W- .0000001
      YMIN=ALAM+ALPHA2*ALOG(ASTOP-W)
      XMINP=1.-EXP(-EXP(YMIN))
      YMAX=ALAM+ALPHA2*ALOG(ALAST-W)
      XMAXP=EXP(-EXP(YMAX))
      IF(IGAM.EQ.3) WRITE(6,6)
      WRITE(6,615) XMINP,XMAXP,BB,AA,W,ITER,DELALF,DELAMB
  615 FORMAT(1H0,2(F10.6,5X),3E15.6,5X,I3,10X,2E15.6)
      IF(IGAM.EQ.3) GO TO 60
      GO TO 501
   60 WRITE(6,2010)
 2010 FORMAT(1H0,T25,*THETA*,T40,*ALPHA*,T55,*GAMMA*,T70,*LOG L*)
      IF(IREFL.NE.0) WRITE(6,2009)
 2009 FORMAT(1H ,T55,*REFLECTED*)
      WRITE(6,2008)
 2008 FORMAT(1H ,)
      DO 300 I=1,K
  300 WRITE(6,2011) I,(PARAM(I,J),J=1,4)
 2011 FORMAT(1H ,10X,I3,5X,3E15.5,E15.6)
      L=1
      IF(K.EQ.1) GO TO 305
      DO 301 J=2,K
      JJ=J-1
      IF (PARAM(J,4).LT.PARAM(JJ,4)) GO TO 301
      L=J
  301 CONTINUE
  305 CONTINUE
      THETA=PARAM(L,1)
      ALPHA=PARAM(L,2)
      GAM=PARAM(L,3)
```

(cont'd on next page)

APPENDIX 9B (cont'd)

```
      ALAMB=EXP(-ALPHA*ALOG(THETA))
      IF(IREFL.EQ.0) GO TO 302
      GAMR=2.*ASTART - GAM
      WRITE(6,303)
  303 FORMAT(1H0,T10,*REFLECTED WEIBULL DISTRIBUTION*)
      WRITE(6,2012) ALAMB,ALPHA,GAMR,THETA,PARAM(L,4)
      GO TO 304
  302 CONTINUE
      WRITE(6,2112)
 2112 FORMAT(1H0,T10,*STANDARD WEIBULL DISTRIBUTION*)
      WRITE(6,2012) ALAMB,ALPHA,GAM,THETA,PARAM(L,4)
 2012 FORMAT(1H0,T10,*MAX LIKELIHOOD ESTIMATES ARE*//T10,*LAMBDA=*,
     CE15.6,10X,*ALPHA=*,E15.6/T10,*GAMMA=*,E15.6,10X,*THETA=*,E15.6/
     CT10,*MAX LOG L=*,E15.6)
  304 CONTINUE
      AX50=ALOG(THETA)-.3665129/ALPHA
      EX=EXP(AX50)
      X50=GAM+EX
      IF(IREFL.EQ.1) X50=GAMR-EX
      WRITE(6,2013) X50
 2013 FORMAT(T10,*L50=*F10.4)
      G1=GMM(1.+1./ALPHA)
      G2=GMM(1.+2./ALPHA)
      UU=THETA*G1
      U1=UU+GAM
      IF(IREFL.EQ.1)U1=GAMR-UU
      VARX=THETA*THETA*(G2-G1*G1)
      SIGMA=SQRT(VARX)
      WRITE(6,2014)U1,SIGMA
 2014 FORMAT(T10,*MEAN=*,F10.4,10X,*STANDARD DEVIATION=*,F10.4)
  600 DO 601 I=1,N
      S(I)=ST(I)-GAM
      IF(S(I).LT.1.E-10) GO TO 601
      V(I)=(S(I)/THETA)**ALPHA
      SLN(I)=ALOG(S(I))
      ARG=V(I)
      IF(ARG.LT.100.) GO TO 602
      Q=0.
      GO TO 603
  602 IF(ARG.LT.1.E-20) GO TO 601
      Q=EXP(-ARG)
  603 P(I)=1.-Q
  601 CONTINUE
      NP=3
      IF(IGAM.EQ.3) NP=2
      CALL COVAR(BF,COV,NP,N,THETA,ALPHA)
      IF(NCR.EQ.0) GO TO 110
      CALL WREL(THETA,ALPHA,GAM,COV,NP)
  110 IF (NPL.EQ.0) GO TO 109
      CALL LP(THETA,ALPHA,GAM,COV,NP,IREFL,ASTART,NPL,PL,NCL,COEF)
      NGM=NGM+1
      LA=0
      GO TO 57
  109 CONTINUE
      GO TO 500
  111 STOP
      END
      SUBROUTINE COVAR(BF,COV,NP,N,THETA,ALPHA)                          COVAR  2
C       FISHER INFORMATION MATRIX IS COMPUTED DIRECTLY AND  INVERTED     COVAR  3
C       TO OBTAIN ASYMPTOTIC COVARIANCE MATRIX.                          COVAR  4
    1 FORMAT(///,T10,*FISHER INFORMATION MATRIX FOR THETA, ALPHA, GAMMA COVAR  5
     C*,/)                                                               COVAR  6
    2 FORMAT(6(1PE15.5))                                                 COVAR  7
    3 FORMAT(1H0,T10,*ASYMPTOTIC COVARIANCE MATRIX FOR THETA, ALPHA, GAMCOVAR  8
     CMA*,/)                                                             COVAR  9
```

(cont'd on next page)

APPENDIX 9B (cont'd)

```
      COMMON/BL1/V,SLN,P,S                                              COVAR 10
      DIMENSION V(150),SLN(150),P(150),S(150),B(3,6),COV(3,6),BF(3,6)
      DIMENSION D(3,3)
       MC=2*NP                                                          COVAR 12
      A=ALPHA/THETA                                                     COVAR 13
      ALT=ALOG(THETA)                                                   COVAR 15
      DO 90 I=1,N                                                       COVAR 16
      SLN(I)=SLN(I)-ALT                                                 COVAR 18
   90 CONTINUE                                                          COVAR 19
       DO 100 I=1,NP                                                    COVAR 20
       DO 100 J=1,MC                                                    COVAR 21
100    B(I,J)=0.                                                        COVAR 22
       DO 110 I=1,N                                                     COVAR 23
      IF(S(I).LE.1.E-10) GO TO 110
      IF(P(I))110,110,201
  201 CONTINUE
       QDP=(1-P(I))/P(I)                                                COVAR 25
       VSQ=V(I)*V(I)                                                    COVAR 26
       B(1,1)=B(1,1)+QDP*VSQ                                            COVAR 27
       B(1,2)=B(1,2)+QDP*VSQ*SLN(I)                                     COVAR 28
       B(2,2)=B(2,2)+QDP*VSQ*SLN(I)*SLN(I)                              COVAR 29
       IF(NP.EQ.2)GO TO 110                                             COVAR 30
      AA=QDP*VSQ*THETA                                                  COVAR 31
      B(1,3)=B(1,3)+AA/S(I)                                             COVAR 32
      B(2,3)=B(2,3)-AA*SLN(I)/S(I)                                      COVAR 33
      B(3,3)=B(3,3)+AA*THETA/(S(I)*S(I))                                COVAR 34
110    CONTINUE                                                         COVAR 35
      B(1,1)=A*A*B(1,1)                                                 COVAR 36
      B(1,2)=-A*B(1,2)                                                  COVAR 37
       B(2,1)=B(1,2)                                                    COVAR 38
       IF(NP.EQ.2)GO TO 120                                             COVAR 39
      B(1,3)=A*A*B(1,3)                                                 COVAR 40
       B(3,1)=B(1,3)                                                    COVAR 41
      B(2,3)=-A*B(2,3)                                                  COVAR 42
       B(3,2)=B(2,3)                                                    COVAR 43
      B(3,3)= A*A*B(3,3)                                                COVAR 44
120    CONTINUE                                                         COVAR 45
       NN=NP+1                                                          COVAR 46
       DO 130 I=1,NP                                                    COVAR 47
       DO 130 J=NN,MC                                                   COVAR 48
      IF(J-NP.EQ.I)B(I,J)=1.                                            COVAR 49
130    CONTINUE                                                         COVAR 50
       WRITE(6,1)                                                       COVAR 51
       DO 140 I=1,NP                                                    COVAR 52
       WRITE(6,2)(B(I,J),J=1,NP)                                        COVAR 53
      DO 140 J=1,NP
  140 BF(I,J)=B(I,J)
      CALL JODIE(B,NP,MC)                                               COVAR 56
       DO 150 I=1,NP                                                    COVAR 57
       DO 150 J=1,NP                                                    COVAR 58
       JJ=J+NP                                                          COVAR 59
150    COV(I,J)=B(I,JJ)                                                 COVAR 60
       WRITE(6,3)                                                       COVAR 61
       DO 160 I=1,NP                                                    COVAR 62
160    WRITE(6,2)(COV(I,J),J=1,NP)                                      COVAR 63
      DO 165 I=1,NP
      DO 165 J=1,NP
      D(I,J)=0.
      DO 165 K=1,NP
      D(I,J) = D(I,J) +BF(I,K)*COV(K,J)
  165 CONTINUE
      WRITE(6,5)
    5 FORMAT(1H0,T10,* PRODUCT OF INFO AND COVARIANCE MATRICES*,/)
      DO 166 I=1,NP
```

(cont'd on next page)

APPENDIX 9B (cont'd)

```
  166 WRITE (6,2) (D(I,J),J=1,NP)
      IF(NP.EQ.2) GO TO 170                                          COVAR 64
      DET=COV(1,1)*COV(2,2)*COV(3,3)+COV(1,2)*COV(2,3)*COV(3,1)+     COVAR 65
     CCOV(1,3)*COV(3,2)*COV(2,1)-COV(3,1)*COV(2,2)*COV(1,3)-COV(3,2)* COVAR 66
     CCOV(2,3)*COV(1,1)-COV(3,3)*COV(1,2)*COV(2,1)                   COVAR 67
      GO TO 180                                                      COVAR 68
  170 DET=COV(1,1)*COV(2,2)-COV(1,2)*COV(2,1)                        COVAR 69
  180 WRITE(6,4) DET                                                 COVAR 70
    4 FORMAT(1H0,10X,*GENERALIZED VARIANCE=*,E15.5)                  COVAR 71
      RETURN                                                         COVAR 72
      END                                                            COVAR 73
      SUBROUTINE WREL(THETA,ALPHA,GAM,COV,NP)                        WREL   2
C       COMPUTES ASYMPTOTIC RELIABILITY ESTIMATES AND CONFIDENCE     WREL   3
C       INTERVALS                                                    WREL   4
    3 FORMAT(/,110,*ASYMPTOTIC RELIABILITY ESTIMATES*//T8,*C*,T20,*REL*,WREL   7
     C T35,*VAR R*,T50,*SIG R*,T65,*C COEF*,T80,*LCL*)               WREL   8
    4 FORMAT(F12.4,3(2X,E13.6),2X,F8.3,2X,E13.6)                     WREL   9
      DIMENSION COV(3,6),C(20),PR(3),COEF(5),Z(5)                    WREL  10
      DIMENSION CR(20)
      COMMON/BLC/NCR,NCL,C,COEF,IREFL,ASTART
C                                                                    WREL  11
      IF(IREFL.EQ.0) GO TO 131
      DO 130 I=1,NCR
      CR(I)=C(I)
      C(I)=2.*ASTART - C(I)
  130 CONTINUE
  131 CONTINUE
      WRITE(6,3)
      A=ALPHA/THETA                                                  WREL  18
      DO 150 I=1,NCR                                                 WREL  19
      IF(C(I)-GAM.GE..1E-05)GO TO 100                                WREL  20
      R=1.0                                                          WREL  21
      IF(IREFL.EQ.0) GO TO 135                                       WREL  22
      R=0.
      C(I)=CR(I)
  135 CONTINUE
      WRITE(6,6) C(I),R                                              WREL  23
    6 FORMAT(F12.4,2X,E13.6,10X,*VARIANCE AND LCL ARE NOT DEFINED*)  WREL  24
      GO TO 150                                                      WREL  25
  100 CONTINUE                                                       WREL  26
      B=(C(I)-GAM)/THETA                                             WREL  27
      V=B**ALPHA                                                     WREL  28
      IF(V.GT. 30.) GO TO 105                                        WREL  29
      R=EXP(-V)                                                      WREL  30
      GO TO 106                                                      WREL  31
  105 R=0.                                                           WREL  32
  106 CONTINUE                                                       WREL  33
      PR(1)=V*R*A                                                    WREL  34
      PR(2)=-V*R*ALOG(B)                                             WREL  35
      PR(3)=PR(1)/B                                                  WREL  36
      VAR=0.                                                         WREL  37
      DO 110 K=1,NP                                                  WREL  38
      DO 110 J=1,NP                                                  WREL  39
  110 VAR=VAR+PR(K)*PR(J)*COV(K,J)                                   WREL  40
      IF(VAR) 124,125,125
  124 IF(IREFL.NE.0) C(I)=CR(I)
      WRITE(6,5) C(I),R,VAR
    5 FORMAT(F12.4,2(2X,E13.6),10X,*VARIANCE IS NEGATIVE*)
      GO TO 150
  125 CONTINUE
      SIG=SQRT(VAR)                                                  WREL  41
      IF(IREFL.EQ.0) GO TO 140
      R=1.-R
      C(I)=CR(I)
```

(cont'd on next page)

APPENDIX 9B (cont'd)

```
  140 CONTINUE
      DO 120 J=1,NCL                                                    WREL  42
      Z(J)=ZNDEV(COEF(J))                                               WREL  43
      CLL=R-Z(J)*SIG                                                    WREL  44
      IF(CLL.LT.0.) CLL=0.                                              WREL  45
  120 WRITE(6,4)C(I),R,VAR,SIG,COEF(J),CLL                              WREL  46
150   CONTINUE                                                          WREL  47
      RETURN                                                            WREL  48
      END                                                               WREL  49
      FUNCTION ZNDEV(P)                                                 ZNDEV  2
C     COMPUTES N(0,1) DEVIATE FOR PROBABILITY P. USING                  ZNDEV  3
C     NEWTON-RAPHSON METHOD OF SOLVING INVERSE RELATION                 ZNDEV  4
      EPS=.005                                                          ZNDEV  5
       B=1/SQRT(2.*3.14159)                                             ZNDEV  6
      Z=0.                                                              ZNDEV  7
90    PHI=.5*(1+ERF(Z*.7071))                                           ZNDEV  8
      DPHIZ=B*EXP(-Z*Z/2.)                                              ZNDEV  9
      DZ=-(PHI-P)/DPHIZ                                                 ZNDEV 10
      Z=Z+DZ                                                            ZNDEV 11
      IF(ABS(DZ).GT.EPS)GO TO 90                                        ZNDEV 12
      ZNDEV=Z                                                           ZNDEV 13
      RETURN                                                            ZNDEV 14
      END                                                               ZNDEV 15
      FUNCTION ERF (Y)                                                  ERF    2
      IF(Y) 3,4,3                                                       ERF    3
    3 CONTINUE                                                          ERF    4
      X = 1.414213562*Y                                                 ERF    5
      AX = ABS(X)                                                       ERF    6
      T = 1.0/(1.0 + .2316419*AX)                                       ERF    7
      D = 0.7978845608*EXP(-X*X/2.0)                                    ERF    8
      ERF = 1.0 - D*T*((((1.330274*T - 1.821256)*T + 1.781478)*T        ERF    9
     1       - 0.3565638)*T + 0.3193815)                                ERF   10
      IF (X) 1,2,2                                                      ERF   11
    1 ERF = -ERF                                                        ERF   12
      GO TO 2                                                           ERF   13
    4 ERF = 0.0                                                         ERF   14
    2 RETURN                                                            ERF   15
      END                                                               ERF   16
      SUBROUTINE JODIE(A,N,M)                                           JODIE  2
      DIMENSION BBX(3,3),C(3,3)
      DIMENSION A(3,6)                                                  JODIE  3
    2 FORMAT(1H0,T20,*NO PIVOT SOLUTION*)
      DO 50 I=1,N
      DO 50 J=1,N
   50 BBX(I,J)=A(I,J)
      CALL INVERT(N,BBX,C)
      DO 20 I=1,N
      DO 20 J=1,N
      JP=N+J
   20 A(I,JP)=C(I,J)
      RETURN                                                            JODIE 33
      END                                                               JODIE 36
      SUBROUTINE INVERT(N,A,AINV)
      DIMENSION X(3)
      DIMENSION A(3,3),UL(3,3),B(3),AINV(3,3)
      CALL DECOMP(N,A,UL)
      DO 1 J=1,N
      DO 2 I=1,N
      B(I)=0.0
      IF(I.EQ.J) B(I)=01.0
    2 CONTINUE
      CALL SOLVE (N,UL,B,X)
      CALL IMPROV(N,A,UL,B,X,DIGIT)
      DO 3 I=1,N
         AINV(I,J)=X(I)
```

(cont'd on next page)

APPENDIX 9B (cont'd)

```
    3 CONTINUE
    1 CONTINUE
      RETURN
      END
      SUBROUTINE DECOMP(NN,A,UL)
      DIMENSION A(3,3),UL(3,3),SCALES(3),IPS(3)
      COMMON IPS
      N=NN
C
C        INITIALIZE IPS,UL,AND SCALES
      DO 5 I=1,N
       IPS(I)=I
      ROWNRM=0.0
      DO 2 J=1,N
      UL(I,J)=A(I,J)
      IF(ROWNRM-ABS(UL(I,J))) 1,2,2
    1 ROWNRM=ABS(UL(I,J))
    2 CONTINUE
      IF(ROWNRM)3,4,3
    3 SCALES(I) = 1.0/ROWNRM
      GO TO 5
    4 CALL SING(1)
      SCALES(I)=0.0
    5 CONTINUE
C
C        GAUSS ELIMINATION WITH PARTIAL PIVOTING
      NM1 =N-1
      DO 17 K=1,NM1
         BIG=0.0
      DO 11 I=K,N
      IP=IPS(I)
       SIZE=ABS(UL(IP,K))*SCALES(IP)
       IF(SIZE-BIG) 11,11,10
   10 BIG=SIZE
      IDXPIV=I
   11 CONTINUE
      IF(BIG) 13,12,13
   12 CALL SING(2)
      GO TO 17
   13 IF(IDXPIV-K) 14,15,14
   14 J=IPS(K)
      IPS(K)=IPS(IDXPIV)
      IPS(IDXPIV)=J
   15 KP=IPS(K)
      PIVOT=UL(KP,K)
      KP1=K+1
      DO 16 I=KP1,N
         IP=IPS(I)
         EM=-UL(IP,K)/PIVOT
         UL(IP,K)=-EM
         DO 16 J=KP1,N
           UL(IP,J)=UL(IP,J)+EM*UL(KP,J)
   16 CONTINUE
   17 CONTINUE
      KP=IPS(N)
      IF(UL(KP,N)) 19,18,19
   18 CALL SING(2)
   19 RETURN
      END
      SUBROUTINE SOLVE(NN,UL,B,X)
       DIMENSION UL(3,3),B(3),X(3),IPS(3)
      COMMON IPS
      N=NN
      NP1=N+1
```

(cont'd on next page)

APPENDIX 9B (cont'd)

```
C
      IP=IPS(I)
     X(I)=B(IP)
     DO 2 I=2,N
      IP=IPS(I)
     IM1=I-1
     SUM=0.0
      DO 1 J=1,IM1
    1      SUM=SUM+UL(IP,J)*X(J)
    2  X(I)=B(IP)-SUM
C
      IP=IPS(N)
      X(N)=X(N)/UL(IP,N)
      DO 4 IBACK=2,N
       I=NP1-IBACK
        IP=IPS(I)
     IP1=I+1
        SUM=0.0
        DO 3 J=IP1,N
    3       SUM=SUM+UL(IP,J)*X(J)
    4   X(I)=(X(I)-SUM)/UL(IP,I)
     RETURN
     END
     SUBROUTINE IMPROV(NN,A,UL,B,X,DIGITS)
     DIMENSION A(3,3),UL(3,3),B(3),X(3),R(3),DX(3)
      N=NN
     EPS=1.0E-15
     ITMAX=30
     XNORM=0.0
      DO 1 I=1,N
    1 XNORM=AMAX1(XNORM,ABS(X(I)))
     IF(XNORM) 3,2,3
    2 DIGITS=-ALOG10(EPS)
      GO TO 10
    3 DO 9 ITER = 1,ITMAX
        DO 5 I=1,N
          SUM=0.0
          DO 4 J=1,N
    4        SUM=SUM+A(I,J)*X(J)
        SUM=B(I)-SUM
    5 R(I) = SUM
        CALL SOLVE (N,UL,R,DX)
         DXNORM=0.0
          DO 6 I=1,N
             T=X(I)
             X(I)=X(I)+DX(I)
             DXNORM=AMAX1(DXNORM,ABS(X(I)-T))
    6    CONTINUE
         IF(ITER-1) 8,7,8
    7      DIGITS=-ALOG10(AMAX1(DXNORM/XNORM,EPS))
    8      IF(DXNORM-EPS*XNORM) 10,10,9
    9  CONTINUE
       CALL SING(3)
   10 RETURN
      END
         SUBROUTINE SING(IWHY)
   11 FORMAT(*0      MATRIX WITH ZERO ROW IN DECOMPOSE*)
   12 FORMAT(1H0,*    SINGULAR MATRIX IN DECOMPOSE. ZERO DIVIDE IN SOLVE
     X*)
   13 FORMAT(1H0,*   NO CONVERGENCE IN IMPROV. MATRIX IS NEARLY SINGULAR
     X*)
      NOUT=3
      GO TO (1,2,3),IWHY
    1 WRITE (6,11)
      GO TO 10
```

(cont'd on next page)

APPENDIX 9B (cont'd)

```
    2 WRITE (6,12)
      GO TO 10
    3 WRITE (6,13)
   10 RETURN
      END
      SUBROUTINE REFLECT (S,U,N,AIMAGE,IGAM,ASTART,ASTOP)
      INTEGER U
      DIMENSION S(150),U(150)
      DO 30 I=1,N
      S(I)=2.*AIMAGE-S(I)
      U(I)=-FLOAT(U(I))+1.001
   30 CONTINUE
      ASTART=2.*AIMAGE-ASTART
      IF(IGAM.EQ.2) ASTOP=2.*AIMAGE-ASTOP
      RETURN
      END
      SUBROUTINE LP(THETA,ALPHA,GAM,COV,NP,IREFL,AIMAGE,NPL,PL,NCL,COEF)
C
C*****PERCENTAGE POINTS,LP, OF STANDARD WEIBULL ARE COMPUTED IF IREFL=0
C*****AND FOR REFLECTED WEIBULL USING REFLECTION POINT AIMAGE IF IREFL
C*****=1.
C
      DIMENSION COV(3,3),PL(30),COEF(5),XP(30),PX(3),SIG(30),Z(5)
      DIMENSION VAR(30)
      WRITE(6,1)
    1 FORMAT(1H1,T10,*WEIBULL QUANTILE ESTIMATES*//T8,*P*,T20,*L(P)*,
     C T35,*SIG LP*,T50,*C COEF*,T65,*LCL*,T80,*UCL*/)
   16 CONTINUE
      PX(3)=1.
      DO 10 I=1,NPL
      IF(PL(I).GT.0.) GO TO 20
      XP(I)=GAM
      IF(IREFL.EQ.1) XP(I)=-9.E+99
      GO TO 24
   20 IF(PL(I).LT.1.) GO TO 25
      XP(I)=9.E+99
      IF(IREFL.EQ.1) XP(I)=2.*AIMAGE-GAM
   24 SIG(I)=0.
      GO TO 10
   25 CONTINUE
      Q=1.-PL(I)
      IF(IREFL.EQ.1) Q=PL(I)
      A=-ALOG(Q)
      PX(1)=A**(1./ALPHA)
      AA=THETA*PX(1)
      XP(I)=GAM+AA
      IF(IREFL.EQ.1) XP(I)=2.*AIMAGE-GAM-AA
      PX(2)=-AA*ALOG(A)/(ALPHA*ALPHA)
      VAR(I)=0.
      DO 30 K=1,NP
      DO 30 J=1,NP
   30 VAR(I)=VAR(I)+PX(K)*PX(J)*COV(K,J)
   10 CONTINUE
      IF(NCL.EQ.0) GO TO 50
      DO 40 I=1,NCL
      ZP=1.-.5*(1.-COEF(I))
      Z(I)=ZNDEV(ZP)
      DO 40 J=1,NPL
      IF(VAR(J)) 60,61,61
   60 WRITE(6,3) PL(J),XP(J),VAR(J)
    3 FORMAT(1H ,F12.4,2X,E13.5,2X,*VAR=*,E13.5)
      GO TO 40
   61 CONTINUE
      SIG(J)=SQRT(VAR(J))
      CLL=XP(J)-SIG(J)*Z(I)
```

(cont'd on next page)

APPENDIX 9B (cont'd)

```
      UCL=XP(J)+SIG(J)*Z(I)
      WRITE(6,2) PL(J),XP(J),SIG(J),COEF(I),CLL,UCL
   40 CONTINUE
    2 FORMAT (1H ,F12.4,2(2X,E13.5),2X,F8.3,2(2X,E13.5))
      RETURN
   50 DO 51 J=1,NPL
   51 WRITE(6,3) PL(J),XP(J),VAR(J)
      RETURN
      END
      FUNCTION GMM(X)
      ETA=X
      ETAF=AMOD(X,1.0)
      IF (ETA) 20,20,22
 20   GMM=0.0
      GO TO 100
 22   IF (ETA-33.0) 25,25,200
 25   GF=((((((.3586834 E-01*ETAF-0.1935278 )*ETAF+0.4821994 )*ETAF-
     1  0.7567041 )*ETAF+0.9182069 )*ETAF-0.8970569 )*ETAF+0.9882059 )*
     2  ETAF-0.5771917 )*ETAF+1.0
      IF (ETA-1.0) 30,32,35
 30   GMM=GF/ETAF
      GO TO 100
 32   GMM=1.0
      GO TO 100
 35   IF (ETA-2.0) 38,32,45
 38   GMM=GF
      GO TO 100
 45   PROD=1.0
      TERM=ETAF+1.0
 52   PROD=PROD*TERM
      IF (TERM-ETA+1.1) 55,60,60
 55   TERM=TERM+1.0
      GO TO 52
 60   GMM=PROD*GF
 100  CONTINUE
      RETURN
 200  ETAM=ETA-1.0
      TWOPI=6.283185
      GMMLOG=ALOG(SQRT(TWOPI))+(ETAM+0.5)*ALOG(ETAM)-ETAM+1.0/(12.0*ETAM
     1)
      GMM=EXP(GMMLOG)
      RETURN
      END
000000000000000000000000
000000000000000000000017
000000000000000000000000
```

CHAPTER 10

THE ROLE OF THE STATISTICIAN IN SCIENTIFIC MODEL BUILDING: ILLUSTRATED FOR THE LIMIT VELOCITY PROBLEM

The role of the statistician in the process of scientific model building is described in this chapter. Usually, the statistician serves either as a consultant who might be able to characterize the model in statistical terms as required, or he sometimes functions as a member of the team that has the overall responsibility for model development. To illustrate the probable role and contributions of the statistician, we select a rather complicated problem—the limit velocity problem—since attempts toward a complete solution may continue into future years, and we illustrate such a challenge to the statistician.

Both the physical and the statistical characterizations of models developed to date are outlined, and the limitations of each are discussed. Possible future avenues of further progress are explored by some analyses of actual data relating to the determination of the limit velocity of target armor.

10-0 LIST OF SYMBOLS

a = constant of proportionality or parameter of a distribution
b = constant, or scale, or shape parameter
D = diameter of penetrator, cm
L = penetrator length, cm
M = mass of penetrator, g
M_R = residual mass, g
M_S = striking mass, g
p = exponent in Lambert model (See Eqs. 10-3 and 10-9.)
T = target (armor plate) thickness, cm or in.
V_L = limit velocity, m/s or ft/s = value of V_S for which $V_R = 0$
V_R = residual velocity of projectile after penetrating armor, m/s or ft/s
V_S = striking velocity of projectile, m/s or ft/s
$V_{0.00}$ = striking velocity for which 0% of the projectiles penetrate the target
$V_{0.10}$ = striking velocity for which 10% of the projectiles penetrate the target
z = $T\sec^{0.75}\theta / D$ = parameter used by Lambert (See Eq. 10-6.)
θ = angle of obliquity at which penetrator strikes the target
ρ = target density, g/cm^3

10-1 INTRODUCTION

During his career, the Army analyst will face a variety of different problem applications of an involved physical nature, and he often will be called upon to help solve these problems or at least to contribute to an "immediate" interim solution. The point is that as a result of many years of data collection and research by several physical scientists, a satisfactory law or model may be available that can be used to interpolate or extrapolate to some specific or perhaps more general conditions. Moreover, the physical model will exhibit the key parameters of interest, usually in proper form, and the results can often be successfully "scaled". The statistician often may contribute to efforts of this type by attempting to deal with and "smooth out" any random variations or "noise", so to speak. Or alternatively, the statistician often may be able to make a quick

fit of the observed data and arrive at an interim solution or "stopgap" model, which could apply with some success for the time being. It is always the "team" effort that pays off better in the long run, of course.

Although this is a handbook and hence a work that would ordinarily present final, useful, and well-tried results, it is believed, nevertheless, that some space should be devoted to the role of the statistician as a team member in an organization primarily engaged in research and development (R&D) in some of the physical, biological, or medical sciences. By devoting a chapter to this particular theme—which is quite important in its own right—it is believed that the role and contributions of the statistician will be enhanced. Moreover, there could be much additional payoff to the organization by joint participation. Thus we have selected a problem of long standing, which we will describe briefly as a physical problem and then will give a summary of it in the statistical sense. Finally, we will give the results of some analyses to date to learn just what the current status of accomplishments is, to point out the limitations, and to indicate just how the physicist-engineer-statistician team might be able to push forward the frontiers of knowledge.

The problem we have chosen involves the penetration of armor, and it is also rather closely related to the problem of sensitivity analyses covered in Chapter 9; the difference is that here we are concerned with a mixture of continuous and discrete distributions while trying to estimate a point of zero percent "responses" or penetrations. In fact, there has been and continues to be the need to determine the residual velocity of a projectile once a piece of armor plate has been hit at any striking velocity and penetrates. In addition, it is highly desirable to estimate the striking velocity that results in a very low or even zero percent chance of penetration. This problem has not been completely solved but, nevertheless, is interesting from both the physical and statistical points of view.

10-2 DESCRIPTION OF THE PHYSICAL AND STATISTICAL ASPECTS OF THE PROBLEM

It is well-known that the more the statistician knows about the physical, engineering, biological, or medical aspects of a problem, the better able he is to make some worthwhile contribution toward a satisfactory solution. Indeed, in many areas of the possible application of statistics, there may already exist some physical laws or models that apply to the problem at hand. Therefore, it becomes mandatory for the statistical contribution to make as much physical sense as possible. In those fields of interest for which no physical models exist, the statistician can often contribute without reference to the physical details. Because it becomes quite important for the statistician to know the physical details of the problem illustrated here, we will present some of the more relevant physical details and parameters involved before proceeding to the statistical description.

10-2.1 BRIEF ACCOUNT OF THE PHYSICAL AND ENGINEERING DETAILS

Penetration of armor studies or the field of penetration mechanics has a very broad and long history, and many capable investigators have contributed in many ways to modeling or describing physically the best forms of laws connecting the key parameters involved. For the case of an armor-piercing (AP) projectile fired at tank armor, one may easily see that the striking velocity V_S of the projectile, the mass M of the projectile, the thickness T of the armor, and the diameter D of the penetrator are all important parameters to the defeat of the armor. In fact, as early as 1886, the Frenchman deMarre formulated a "dimensionally awkward" equation that involves the so-called "limit velocity" of the armor plate. The deMarre equation is

$$MV_L^2/D^3 = aT^{1.4}/D^{1.5} \tag{10-1}$$

where

V_L = limit velocity
a = constant of proportionality.

Grabarek (Ref. 1) and others have defined the limit velocity V_L as the lowest striking velocity of a projectile required for a complete penetration of a target. "Complete penetration means that the penetrator exits the rear face of the (armor) target. V_L is determined by test firings wherein the striking velocity, V_S, of the

penetrator and its residual velocity, V_R, are measured." These measurements are usually made with the assistance of flash radiography. "Generally, V_L is determined to within ±5 m/s." The reader will note, however, that there are some very real problems with this definition, i.e., finding the lowest "striking" velocity of projectiles that "penetrate" the armor plate seems to ignore the fact that many of the shots for low V_S do not even "penetrate"!* (We will, therefore, use a somewhat different definition in the statistical treatment in par. 10-2.2.) In any event, the deMarre equation does give a physical relationship between important parameters having a bearing on the "defeat" of the armor plate for normal (perpendicular) incidence attack. Note that the deMarre law is really expressed in terms of a measure of projectile energy, penetrator diameter, and plate thickness.

Although the deMarre equation (Eq. 10-1) may be informative, it is best to refer also to a graph to see more clearly the actual physical situation. On Fig. 10-1 we have plotted the residual velocity of a long-rod penetrator emerging from a piece of tested armor plate versus the striking velocity of the projectile. Fig. 10-1 is the same as Fig. 6-1, and some discussion of this particular problem has also been given in par. 6-3.2, in which a straight line of the square of the residual velocity versus the square of the striking velocity has been fitted to the data as indicated by the equation of the graph. The graph of V_R versus V_S is not linear, but rather it is very sharply curved for the lower striking velocities. The reader will note again that at the very high striking velocities the residual velocities are nearly equal to or approach the corresponding striking velocities, and the slope of the curve becomes unity (45 deg). On the other hand, as the striking velocity decreases, the residual velocity becomes much, much less than the striking velocity, the curve drops very sharply, and at about 2500 ft/s striking speed some or many of the projectiles will not even penetrate the plate. Moreover, the slope of the curve becomes vertical (infinite). The terminal ballistician's definition of the critical or limit velocity is apparently the lowest striking velocity of the rounds that penetrate the plate and mention nothing of the nonpenetrating rounds! We will, however, take the nonpenetrating projectiles into consideration in par. 10-2.2.

The deMarre equation (Eq. 10-1) represents a relationship of some, but perhaps not all, of the key parameters, and the physical scientist does not know the exact or true law. Rather, he is looking for a physically meaningful model except for the random or residual scatter about the law, so to speak. On Fig. 10-1 it is seen that the equation relating the squares of the residual and striking velocities fits the data fairly well, whereas the deMarre equation (Eq. 10-1) uses only the limit velocity V_L that appears to be about 2500 ft/s. (The limit velocity is predicted from the equation on Fig. 10-1 to be the square root of 7,271,000, or approximately 2477 ft/s.)

With reference to a search for the best physical law, the deMarre equation has been somewhat generalized, as indicated by Lambert (Ref. 2), to the form

$$MV_L^2/D^3 = a(T/D)^b \tag{10-2}$$

where a and b are constants that may be determined. Note that Eq. 10-2 may be linearized by taking logarithms of both sides, and indeed a linear least squares fit could be found for the data.

On Figs. 2, 3, and 4 of his Ballistics Research Laboratory (BRL) Memo Report No. 2134, Grabarek (Ref. 1) indicates a fairly good linear relation between the left-hand side (LHS) of Eq. 10-2 and the quantity $T\sec\theta/D$, in which the angle θ is the striking angle or the obliquity of the projectile against the armor. Fig. 10-2 reproduces Fig. 4 of Ref. 1, which shows that a rather simple law and linear relationship have been found for the parameters involved although the residual velocity cannot be predicted from any striking velocity of a penetrator. This brings us to our objectives, which may be stated more clearly now, concerning the problem. We would like to estimate the critical or limit velocity, which is obviously of considerable interest in projectile and armor plate design, and we would also like to know just how good our estimate is. Perhaps the latter could be determined by being able to place confidence bounds about the true, but unknown, limit velocity. Also we would like to be able to estimate the residual velocity of a penetrator "precisely and accurately", given the striking velocity of the projectile. Hopefully, moreover, we should find a "physical" relationship that can be

*For this definition, it is seen that V_R unfortunately will depend on the number of rounds fired (sample size)!

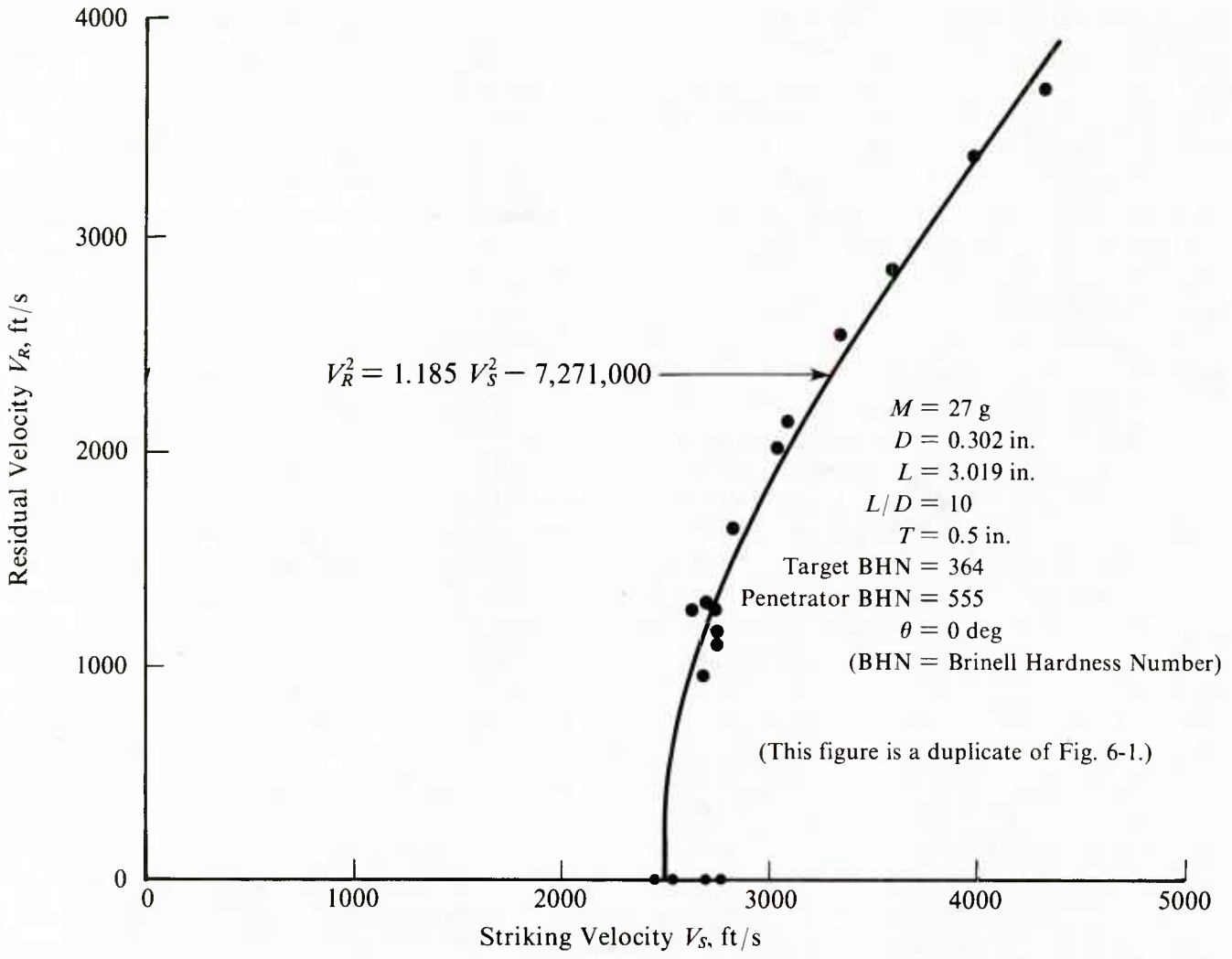


Figure 10-1. Plot of Typical Residual and Striking Velocities for a Penetrator Against Armor

used more or less as a "general law" from which to make predictions. Clearly, these are very demanding objectives although we could add that we also want as simple a law as possible! We should add that it cannot be expected that a suitable law could be found that would include all of the key variables or parameters of interest and still be of unquestionable merit.

To continue the discussion, it appears that some rather intense interest has developed in connection with a proposed law or fit by Lambert and Jonas (Ref. 3, 1976). Their model takes the form

$$V_R = a(V_S^p - V_L^p)^{1/p} \tag{10-3}$$

where a and p are determinable constants, and the equation is to be used only for striking velocities exceeding the limit velocity. In Ref. 2 Lambert extended the work of Ref. 3 to include equations for the determination of the constants a and p. We will discuss the equations after citing further pertinent references concerning the work of other key investigators.

In an earlier report Bethe (Ref. 4) used elasticity theory to analyze the action of the armor plate in stopping penetrators. In fact, he determined that limit energy is proportional to the quantity TD^2 and thus concluded that in Eq. 10-2 the exponent $b = 1$ should be the case. Zener and Holloman (Ref. 5) further studied the mechanism of armor penetration, and during World War II (1943) H. H. Robertson (Ref. 6) of the National Defense Research Council made several profound contributions to the penetration mechanics theory for attacking armor, taking into account the pioneering work of Poncelet (1840). Poncelet hypothesized that the

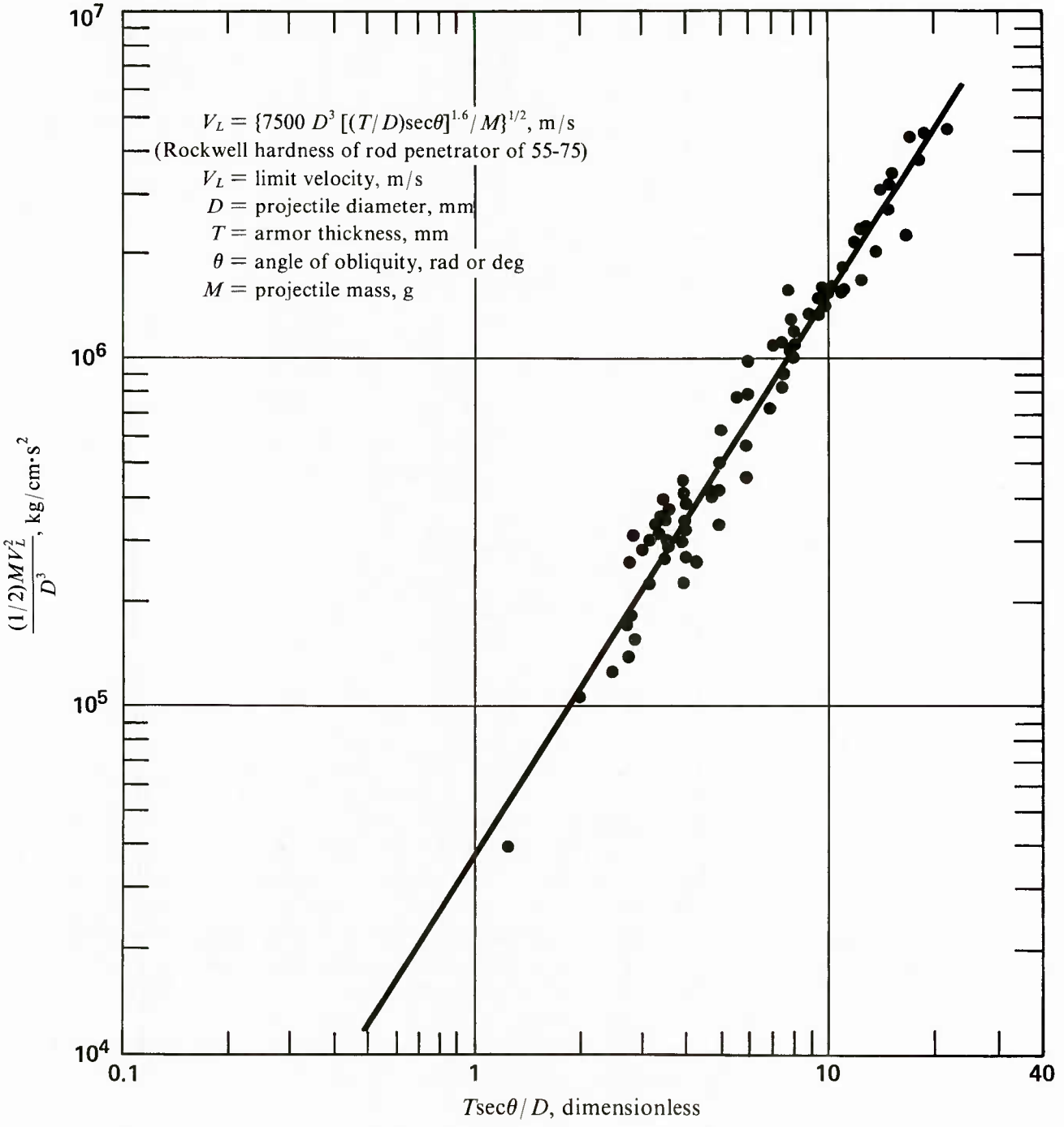


Figure 10-2. Linear Relationship Between Specific Impact Energy and Scaled Armor Thickness (Ref. 1)

resistance encountered by a penetrator passing through a plate is a linear function of the square of the velocity of the penetrator. Taub and Curtis, in an addendum to one of Robertson's reports (Ref. 6), discuss the limit velocity formulations inspired by the Poncelet and Bethe theories and consider the Bethe theory to be valid while the penetrator is in the main body of the plate, but the mechanism of failure changes to a petaling-type situation near the back of the plate. Thus Taub and Curtis (Ref. 6 addendum) derive the law

$$MV_L^2/D^3 = a(T/D + b) \tag{10-4}$$

for which a and b are constants. The Taub and Curtis development of Eq. 10-4 supposes that the ratio of

backface thickness, where petaling prevails, to penetrator diameter is constant, and the constant b then is a quadratic function of that constant value. (For an extensive and interesting account of the early approaches to determination of limit velocity, the reader should refer to the work of Curtis, Ref. 7.)

Based on a study of Refs. 1-7 and much data gathered for armor penetration investigations over the years, Lambert (Ref. 2) has advanced the following equations for estimating the limit or critical velocity V_L and the residual velocity V_R for a plate of rolled homogeneous armor

$$V_L = 4000(L/D)^{0.15}[(z + e^{-z} - 1)D^3/M]^{1/2}, \text{ m/s} \tag{10-5}$$

where

$$z = (T/D)\sec^{0.75}\theta, \text{ dimensionless} \tag{10-6}$$

and the residual velocity is

$$V_R = a(V_S^p - V_L^p)^{1/p}, \text{ m/s} \tag{10-7}$$

where

$$a = M/(M + \rho\pi D^3 z/12), \text{ g} \tag{10-8}$$

$$p = 2 + z/3, \text{ dimensionless} \tag{10-9}$$

where

T = armor thickness, cm
V_S = striking velocity of projectile, m/s
L = penetrator length, cm
D = diameter of penetrator, cm
M = mass of penetrator, g
θ = angle of obliquity at which penetrator strikes the target, rad or deg
ρ = target density, $g/cm^3 = 7.8\ g/cm^3$ for rolled homogeneous armor.

Therefore, the Lambert equations (Ref. 2) predict both the limit velocity of the plate in terms of the projectile length, diameter, mass, plate thickness, angle of obliquity, and the residual velocity of a projectile penetrating the plate in terms of the parameters a and p of Eqs. 10-8 and 10-9 using Eq. 10-7.

A very important consideration is that Eqs. 10-5 through 10-9 use the key physical parameters or constants of the projectile and armor plate and, hopefully, describe a rather general region of application for any prediction purposes. Even for predicting the residual velocity of the projectile emerging from the armor plate after penetration, the "law" (Eq. 10-7) gives a relationship among the striking velocity, the residual velocity, and the desired limit velocity in terms of the parameters a and p. We note that a and p are functions of the projectile mass, diameter, plate thickness and plate density, and the striking angle of obliquity. It must be added, however, that Eq. 10-7 should certainly be suspect! To begin with, the power or exponent p will ordinarily be fractional, so could such a law represent a meaningful "physical" application? In fact, is not the value of p "dimensionally awkward"?

Another and perhaps more pertinent comment on Eq. 10-7 is that it contains the limit velocity V_L as somewhat of a "nuisance" parameter because V_L is required to predict V_R when V_S is given or known. On the other hand, for example, the equation on Fig. 10-1 gives V_R as a function of V_S, and for $V_R = 0$ the striking velocity V_S then becomes equal to the limit velocity V_L without the need for V_L as a parameter. Moreover, confidence bounds on V_L using Eq. 10-7 are most difficult to obtain!

In any event, we have more or less described the state of the art in physical terms for a very involved problem, but it does not appear that a completely satisfactory solution is near. Indeed, it would seem that a considerable amount of additional research needs to be done to obtain a continuing and quite general physical law. Perhaps the statistician could contribute here by "ironing" out the "noise", so to speak. However, it certainly seems true that the physical and engineering aspects of the problem are not completely in hand, so that we might logically ask, "What can the statistician contribute?".

10-2.2 THE STATISTICAL APPROACH

One statistical approach is for the statistician to help the physical scientist to "filter or average out the noise" for the physically formulated law, especially if the law is otherwise completely satisfactory. In fact, exactly this is often done, and the job is so finished. On the other hand, this condition does not hold for the present endeavor or application because more investigation seems warranted and we can indeed formulate a most interesting statistical concept of population mixtures.

Although it is not possible to fire projectiles at a target so that they will have the same striking velocity (due to the random muzzle velocities of the weapon), let us visualize that we could accomplish just this, beginning with some high level of striking velocity. Then the reader should understand that for a constant striking velocity there would be a (probability) distribution of residual velocities for the penetrating projectiles. At the high striking velocities, all, or practically all, of the projectiles would perforate the armor plate. As the striking velocity of the projectiles is reduced, we would approach the situation for which not all projectiles penetrate the plate. Moreover, as we decrease the striking velocity, we can see that we would go from the condition in which 99% of the projectiles penetrate on through the condition in which 95% penetrate, beginning somewhere up above the knee of the curve on Fig. 10-1, perhaps at about 3000 ft/s. For the 1% or 5% not penetrating, the residual velocities are all zero. Thus suddenly we have run into a mixture of continuous and discrete probability distributions. In fact, for each level of striking velocity below the "knee" of the curve of Fig. 10-1, there exists a binomial population with a parameter equal to the fraction of projectiles not penetrating the plate (or the complement of that fraction, if we prefer), and of the fraction of the projectiles perforating the plate, we have a distribution of residual velocities.

As the striking velocity is decreased farther, we soon reach the median or 50% point for some striking velocity—which was discussed in Chapter 9, using only the discrete variable of either a penetration or a nonpenetration. (Note here, however, that the median or $V_{0.50}$ striking velocity is not easy to estimate either from the graph of Fig. 10-1 or from the mixture of continuous- and binomial-type distributions. Indeed, one would have to fire many rounds to estimate the median striking velocity—see Chapter 9.)

As the striking velocity is decreased, it is easily seen that the proportion or fraction of rounds not penetrating the armor will increase, ultimately to 100%, after we pass through the $V_{0.10}$, $V_{0.05}$, $V_{0.01}$, etc., points for the striking velocity. We will then reach the "limit" velocity V_L as defined in par. 10-2.1 by Grabarek (Ref. 1), and finally it may be seen that the "limit" velocity for zero percent penetrations $V_{0.00}$, as we may call it, will be attained. (We have indicated that the limit velocity as defined by the terminal ballistician in par. 10-2.1 may be different from the striking velocity for zero percent penetrations, perhaps due especially to the "physical" definition of limit velocity, which considers only the penetrating rounds. Note in this connection on Fig. 10-1 that three rounds in that test did not penetrate at a bit above the limit velocity, and one round did not penetrate just below the limit velocity. In fact, just above the critical velocity there would be practically no perforations. This example should serve to be a very convincing case of illustrating the experimental need for a huge number of rounds or observations!)

In summary, we have an interesting problem that is both physical and statistical. Moreover, it is also a case for which both the physical and statistical analyses are needed. For example, it does not seem very fruitful to attempt to estimate key parameters by treating the problem only as a statistical problem of some mixture of continuous- and binomial-type populations. In fact, it is very difficult to conduct the needed experiments that way, and the binomial populations change so fast around and below the knee of the curve that efficient sampling may not be possible. If it is desired to estimate the median or striking velocity for 50% perforations, the statistical analysis of Chapter 9 may be needed. However, to estimate the limit velocity by statistical methods may turn out to be very costly in sample size, whereas with some worthwhile physical theory available it could be easier to determine the limit velocity accurately enough for projectile and plate design parameters. It seems, as a matter of fact, that it may be appropriate to determine $V_{0.50}$ by using the methods of Chapter 9 and to estimate the limit velocity or $V_{0.00}$ by a fitted curve as in Fig. 10-1. In any event, it appears that we are faced with a problem for which any completely accurate description of the statistical distributions may not be really needed. Rather, the terminal ballistician will be concerned primarily with predicting the limit velocity and the residual velocity for any striking conditions.

At this point, it seems highly desirable to reemphasize that the scope of any appropriate analysis should include not only estimation of limit velocity, but also should pay vital attention to the determination of just how good that estimate is. Thus it would be quite important to be able to place confidence bounds about the true unknown limit or critical velocity. Consequently, we will keep this point in mind, especially for the statistical analysis.

How then can the statistician contribute? To begin with, this has already been done in par. 6-3.2, in which we found the linear regression of the residual velocity squared on the square of the striking velocity, as is plotted on Fig. 10-1. In this connection it was assumed that the residual velocity squared was linearly related to the striking velocity squared, and the equation established is, as shown on Fig. 10-1,

$$V_R = (1.185V_S^2 - 7{,}271{,}000)^{1/2}, \text{ ft/s} \tag{10-10}$$

which is a very simple relation between the residual and striking velocities of the long-rod penetrator data of par. 6-3.2. We should note for this linear fit that only the striking and residual velocities were used to determine Eq. 10-10 by the method of least squares. Thus the mass of the projectile (27 g) and the thickness of the armor plate (0.5 in.) were not used, nor was the diameter of the long-rod penetrator or any metallurgical characteristics of the plate and projectile. The generality of application of Eq. 10-10 would therefore be questionable although it does apply to this particular projectile-armor combination. Eq. 10-10 does make some physical sense, nevertheless—it cannot only be used to predict the residual velocity for any striking velocity of the 27-g penetrator, but setting the residual velocity equal to zero, we obtain the critical velocity of 2477 ft/s. Also we may easily determine confidence limits about the true unknown critical velocity. The 95% confidence limits about V_L are found in Chapter 6 and Ref. 8 (p. 27) to be 2413-2539 ft/s or a width of 126 ft/s if Eq. 10-10 is used. Therefore, we have the additional advantage of confidence bounds if the statistical fit is determined.

Since the reader is likely thinking of it, we should remark that a direct least squares fit of V_R on V_S could have been determined although we desired to obtain an approximate linear fit so that confidence bounds could be placed easily about the true critical or limit velocity. (The fit so obtained would represent the branch of a hyperbola.)

An appropriate question at this point would be whether a better least squares fit could not be obtained statistically so that we could improve on the width of the confidence bounds, or 126 ft/s. This can, in fact, be done by including the mass of the penetrator before and after perforation of the armor or, in particular, by determining the linear regression of the residual energy on the striking energy. In other words, given the "punching" energy of the projectile, which uses the full weight of the penetrator and its striking velocity, one can predict the residual energy from a linear relation. If this predicted residual energy is divided by one-half the remaining mass (and hence a random amount) of the projectile after penetration, one obtains the residual velocity squared, and the square root gives the desired residual velocity. Precisely this has been done in Ref. 8 (the residual mass data is given in Table II), and the least squares equation is then found to be

$$V_R = (1.457V_S^2 - 9{,}335{,}540)^{1/2}, \text{ ft/s.}^* \tag{10-11}$$

By putting $V_R = 0$ in Eq. 10-11, one finds that the striking velocity or the limit velocity becomes $V_L = 2531$ ft/s versus the 2477 ft/s obtained by the use of Eq. 10-10. Moreover, the 95% confidence bounds on the true limit velocity now become (2497 − 2565) ft/s or only a width of 68 ft/s for the regression of residual energy on striking energy. This amounts to a decrease of 126 − 68 = 58 ft/s in the width of the confidence bound. Hence we should conclude that a better fit is obtained by using the residual energy versus the striking energy since we can predict the limit velocity and the residual velocities with much greater precision.

Eq. 10-11 accounts for both the projectile mass and its striking velocity although it is very difficult to "get a handle" on the residual mass of the projectile after penetration because some random amount up to a third of the projectile weight will "wear away" in the perforation process. Nevertheless, in considering the residual energy versus the striking energy, we do clearly have a physical law relationship in Eq. 10-11, and the

*Note that Eq. 10-11 relating energies is somewhat different from Eq. 10-10. See par. 6-3.2 also.

prediction is precise—something that is neither directly nor easily obtainable with the use of the physical laws of Eqs. 10-5 through 10-9 developed in Ref. 2. The statistical regression of residual energy on striking energy does indeed make a very simple linear model from which to place confidence bounds about the true unknown limit or critical velocity—Eq. 10-11. In this connection, it might be worthwhile to investigate the use of residual versus striking energy for a variety of plate thicknesses and projectile diameters (and lengths) to see whether some scaling effect could be easily incorporated into such a law. At least, this approach may be at least as promising as trying to work a statistical fit into the physical laws of Eqs. 10-5 through 10-9.

With reference to a quantitative comparison of Eqs. 10-5 through 10-9 and the regression of Eq. 10-11 at this point, in Ref. 2 Lambert states "This model for limit velocity adapts remarkably well to our 200-item limit velocity data base. The root-mean-square error associated with the fit of model to data is 65 m/s; the average absolute error (difference between experimental value and model estimate) is 52 m/s and the average absolute percentage error is 4.4%.". Thus from a rather large data base and for a wide range of conditions, Eq. 10-5 of Lambert (Ref. 2) appears to predict the limit velocity with a standard error of approximately $65(39.37/12) = 213$ ft/s, whereas the equivalent standard error for Eq. 10-11 is less than 30 ft/s for the single sample fit involving only the striking velocity and masses. Hence while it cannot be expected that a precise physical law can easily be found to fit such a wide variety of conditions, the statistical analysis would nevertheless indicate that since such a good fit can be obtained by using only two key parameters, perhaps much more needs to be investigated from the physics of the problem. Indeed, a team effort involving both the terminal ballistician and the statistician could well be in order because there may still be some missing but important parameters that should be considered. This brief analysis should provide rather convincing evidence that the terminal ballistician should not be completely satisfied with the ubiquity of application of Eqs. 10-5 through 10-9.

As a result of this statistical characterization and analysis, it should become clear to the terminal ballistician that some very low level of probability of penetration should be used as protection and not a limit velocity dependent on the number of rounds fired.

Although so far for the statistical analysis we have described the limit velocity problem as a mixture of continuous and binomial distributions, another way to examine the overall representation or characterization is to hypothesize that for some (low) striking velocity the chance of a penetration or perforation will start from zero and increase as the striking velocity increases. For some rather high striking velocity, the percent of armor penetrations will approach one hundred. Thus it could be hypothesized that a cumulative frequency distribution may be fitted to the data. Of course, there may be some failures to penetrate, which would result in the corresponding residual velocity being zero, but there would also be residual velocities matching the corresponding striking velocities at the higher levels. This characterization brings up the question of which distribution should be fitted. It could be exponential for simplicity, or normal, etc., but, for the variety of possible shapes that may be encountered, the Weibull distribution seems quite valid indeed. This is precisely the assumption of Clark, Crow, and Sperrazza in their statistical treatment of the limit velocity problem as covered in Ref. 9. A special case of the Weibull fit is the exponential, which has been studied, for example, by Johnson, Collins, and Kindred (Ref. 10), who consider the exponential model

$$V_R = V_S - V_L \exp[b(1 - V_S/V_L)] \tag{10-12}$$

where

b = constant.

(Actually, the adjustment has to be made so that b and V_L are both determined in the fitting process.)

For further (not altogether statistical) suggestions on fitting limit velocity type data, the reader is referred to the hyperbolic fit of Bruchey (Ref. 11), i.e.,

$$V_R = aV_S^2 + b \tag{10-13}$$

and other studies on the subject by Kokinakis and Essig (Ref. 12) and by Morfogenis (Ref. 13). All of this background material will be of interest especially to those who desire to continue research on the subject.

The Weibull model suggested by Clark, Crow, and Sperrazza in Ref. 9 takes the form

$$V_R = V_S\{1 - \exp[-a(V_S - V_L)^b]\} \tag{10-14}$$

where a and b are the scale and shape parameters, respectively, to be determined, as is also the start of frequency or absolute zero probability point V_L the limit velocity.

The Weibull model of Eq. 10-14 can always be linearized by dividing through by V_S, transposing the one, and taking logarithms twice. However, the limit velocity V_L is still a very troublesome nuisance parameter, and the least squares adjustment is best made with the aid of a computer. Appendix A of Ref. 9 gives a nonlinear programming algorithm for fitting Eq. 10-14 by the method of least squares; this is for the three-parameter Weibull model. Also the computer program uses all of the striking velocities for which the residual velocities are zero, as is the case for the linear regression of residual energy on striking energy in Eq. 10-11. Note, that in making the least squares adjustment, the limit velocity V_L is found along with the shape and scale parameters in the process.

In Ref. 9 the Weibull model of Eq. 10-14 and the hyperbolic model of Eq. 10-13 are compared using eight sets of penetration data. In four of the eight cases, the Weibull model gave variances of residuals smaller than the hyperbolic model. A limitation of both fits, however, is that confidence bounds on the limit velocity are not readily obtainable, but they are for the simple linear regression of the residual energy on the striking energy.

The computer algorithm of Ref. 9 for the Weibull model has been used for the data of Table 6-2, or Table II of Ref. 8, to estimate the critical velocity and the shape and scale parameters of the three-parameter Weibull fit. The established relation between the residual and striking velocities is

$$V_R = V_S\{1 - \exp[-0.02867(V_S - 2512.5)^{0.5777}]\}, \text{ ft/s} \tag{10-15}$$

so that the Weibull fit is subexponential with a shape parameter of 0.58, and the critical velocity is estimated as 2512.5 ft/s, as compared, for example, to the value of 2531 ft/s estimated by using the linear regression of residual on striking energy. There is another way to look at a comparison of the two fits, and that is by comparing the standard deviations of the residuals, i.e., the "root-mean-square of the observed minus the predicted values of residual velocities based on Eq. 10-15". For the Weibull model of Eq. 10-15 the standard deviation of residual minus fitted velocities is estimated to be about 124 ft/s. On the other hand, for the simple linear regression of residual on striking energy, the corresponding standard error of residuals is estimated to be only about 60 ft/s. We emphasize in this connection that for the physical fit of a linear relation of residual on striking energy, we used the observed masses of the penetrators after perforation of the armor. Of course, there would always be some difficulty in the determination of these masses. Nevertheless, since the linear regression of residual versus striking energy gives a standard deviation of residuals about half that of the Weibull fit, this again raises the question concerning whether or not the physical fit is superior to any statistical model. Both points of view have provided a considerable amount of insight.

Examination of Eq. 10-15 will reveal that we actually fitted the ratio V_R/V_S, a quantity less than unity, to the cumulative frequency distribution assumed to be Weibull in form. Thus many readers may recall that in fitting life-length data with a Weibull model, we deal with only one set of ordered observed sample values. Therefore, in this connection one can see that the striking velocities could be ordered and only these could be used to fit the assumed Weibull model. Also one might consider truncating those striking velocities for which the residual velocities are equal to zero and then ordering the remaining striking velocities of the total sample. In fact, many Weibull data fits are made from available theory in this manner. One could use the methods outlined in Chapter 21 of Ref. 14 and fit a three-parameter Weibull model (by adjusting values of the location parameter to give minimum variance of residuals) and compare such results with the fit of Eq. 10-15. This would give another statistical prediction of the limit velocity.

Having presented both the physical and statistical points of view for the limit velocity type of problem, we will now bring the results together and comment further on this type of effort.

10-3 DISCUSSION OF THE STATE OF THE ART OF PHYSICAL AND STATISTICAL ESTIMATION OF LIMIT VELOCITY

Perhaps an appropriate summary of the current state of the art of the methods of estimation of critical or limit velocities can best be described and compared by bringing the results together as briefly summarized in Table 10-1. We believe that our key points of discussion can be properly highlighted by displaying four methods of estimation of limit velocity. These are (1) the Grabarek linearization approach of Ref. 1 and Fig. 10-2, (2) the approach of Lambert (Ref. 2) that uses Eq. 10-5, (3) the Weibull fit of Clark, Crow, and Sperrazza (Ref. 9), and finally (4) the linear regression of the residual energy on striking energy of Ref. 8. We use the data of Table 6-2 here.

Before any detailed discussion of Table 10-1, which is very illuminating and revealing of the status of the limit velocity problem as of 1979, we should provide some orientation. Initially, we desire to develop a model or, in fact, the correct model that uses all of the key parameters to predict the limit velocity of any projectile-armor plate combination. This means we must use the diameter D of the penetrator, the mass M of the penetrator, the length L of the projectile, some measure of the metallurgical properties of the penetrator including its hardness, perhaps the shape of the nose of the projectile, the thickness T of the armor plate, the angle θ of striking obliquity, some constants or parameters describing the metallurgical properties of the plate and its hardness (very likely the density of the plate and that of the penetrator), and any constants that may appear in an "empirical" relationship between the numerous parameters—to mention some of the parameters we think will be "key" variables. If, in addition to the estimation of the limit velocity, we would like to determine the residual velocity, we would expect to be given the striking velocity also. Therefore, we could state that we may need to fit 10-12 parameters into our "model". However, if we use a lesser number of parameters, they should account for the others or at least leave very little "noise" or random, unaccounted for variation—i.e., variance of residuals.

TABLE 10-1

COMPARISON OF LIMIT VELOCITY ESTIMATION METHODS

Method	Limit Velocity Estimated, ft/s	Confidence Bounds, 95%	Comments
Grabarek Ref. 1 (Fig. 10-2)	2526	Could be obtained*	Uses M, D, T, and θ to determine V_L
Lambert Ref. 2 (Eq. 10-5)	2397	Very difficult to obtain	Uses M, D, T, θ, and L to determine V_L
Weibull Ref. 9 (Eq. 10-14)	2513	Approximate bounds available	Uses only V_S and V_R to determine V_L
Residual versus Striking Energy (Eq. 10-11)	2531	2497-2565 ft/s Easily and naturally obtained	Uses only V_S, V_R, M_S, and M_R to determine V_L

*Since the Grabarek method of Ref. 1 is a linearization, the determination of confidence bounds is really quite straightforward.

Now let us turn to an examination of Table 10-1. Initially, we see immediately that none of the four models uses all of the desired parameters or variables. The Lambert model (Ref. 2) uses five parameters (more than any other model), and the Grabarek model uses four of the "thought-to-be" key parameters, whereas the two statistical fits may omit too many important or key variables of interest. For example, the Weibull fit does not use penetrator mass, penetrator diameter, target thickness, any metallurgical properties, hardness, or the

angle of obliquity. The residual versus striking energy linear regression does not use penetrator diameter, target thickness, any metallurgical properties, hardness, or angle of obliquity either although the concept of "punching energy" may clearly be in the right direction. (The angle of obliquity may be considered to be taken care of by an equivalent thickness of the target armor plate; however, the other parameters must be taken account of, obviously.) The two statistical fits, therefore, seem rather simplistic and are, therefore, only a start toward any completely acceptable solution to the limit velocity problem. This is not to say, however, that the statistical fits would not be useful for a given set of fixed conditions.

The Lambert model appears to be of considerable interest, especially since it considers the five key variables. Nevertheless, the fractional exponent p does seem to deviate from any completely acceptable physical model, and Eq. 10-7 does not extrapolate to a residual velocity of zero, which indicates that it may be somewhat questionable. There is some evidence also that the Lambert model to determine V_L may underestimate the limit velocity by about 120 ft/s (Table 10-1), it being this much lower than the others. While it is realized that this particular calculation is an isolated one, it does seem clear that the Lambert model needs some improvement. For example, it does not acknowledge the metallurgical properties of the penetrator and its hardness, nor does it account for the shape of the penetrator nose—if that is important. Note also that nonlinear least squares fits would have to be made for the model and that confidence bounds on the limit velocity are not easy to achieve. Otherwise, it does seem that most of the highly key parameters are accounted for in the Lambert model.

The Grabarek model (Ref. 1 and Fig. 10-2) apparently does not account for penetrator length, sectional density, the Brinell hardness number (BHN) of either the penetrator or target, or the projectile shape (if important)—to mention some additional parameters. The effect of including these additional parameters in a model of the fitted line or curve, therefore, is not known. Nevertheless, Fig. 10-2 indeed indicates that the linear relationship is rather well-established over quite a range of parameters. In this connection, does it mean that the fitted law is correct? One should examine the residuals about the fitted line to see whether the larger ones could be physically considered and hence improve upon the selected model. In fact, some of the deviations about the fitted line appear rather large in magnitude, which indicates the need perhaps for further investigation. Perhaps the statistician could make a contribution by using the methods of Chapter 3 to detect the outlying residuals, or he could also perform some least squares adjustments to fit the best law or model, as in Fig. 10-2.

For a more complete account of the dynamics of ballistic impact, the reader should study Ref. 15. This handbook gives wide coverage of many important topics in terminal ballistics, and Chapter 4, especially pars. 4-2 and 4-3, discusses additional details of some of the subjects of this chapter from a different point of view.

Two other references that might be of interest are Refs. 16 and 17. Ref. 16 discusses a regression approach that includes many parameters of interest from which to predict, and Ref. 17 is a handbook of equations and computer programs for kinetic penetrators, including fragments.

Much additional work seems necessary insofar as the limit velocity problem is concerned, and perhaps it will take years to settle the remaining important issues. A straightforward, "textbook" statistical approach to the limit velocity problem may leave much to be desired because it would ignore too many important physical parameters, and the need to develop a good, entirely acceptable physical model will require some special nonstatistical expertise. Nevertheless, there does seem to be quite an important role for the statistician; he is very much needed in the team effort. In fact, we believe that a team effort involving both the terminal ballistician and the statistician will be necessary to make any further significant progress.

10-4 SUMMARY

To illustrate the role of the statistician as part of any team effort toward model building, we have selected a rather involved, continuing, and as yet unsolved problem in terminal ballistics—namely, the limit or critical velocity problem. We have oulined briefly the physical or terminal ballistic accomplishments to date, and we have given an account of some statistical attainments. In this connection, it becomes unmistakably clear that real progress toward a lasting solution will depend on a team effort involving both terminal ballisticians and statisticians. Such a team effort is required for many current endeavors in Army research, development, testing, and elsewhere as well, we believe.

REFERENCES

1. C. L. Grabarek, *Penetration of Armor by Steel and High Density Penetrators*, US Army Ballistic Research Laboratories Memorandum Report No. 2134, Aberdeen Proving Ground, MD, 1971.
2. J. P. Lambert, *A Residual Velocity Predictive Model for Long-Rod Penetrators*, US Army Ballistic Research Laboratories Memorandum Report ARBRL-MR-02828, Aberdeen Proving Ground, MD, April 1978.
3. J. P. Lambert and G. H. Jonas, *Towards Standardization in Terminal Ballistics Testing: Velocity Representation*, US Army Ballistic Research Laboratories Report No. 1852, Aberdeen Proving Ground, MD, 1976.
4. H. A. Bethe, *Attempt of a Theory of Armor Penetration*, Ordnance Laboratory, Frankford Arsenal, Philadelphia, PA, 1941.
5. C. Zener and J. Holloman, *Mechanism of Armor Penetration, First Partial Report*, Watertown Arsenal Report 710/454, Watertown, MA, 1942.
6. H. P. Robertson, *The Mechanics of Armor Perforation, I. Residual Velocity*, National Defense Research Council, Armor and Ordnance Report A-227, Office of Scientific Research and Development, OSRD No. 2043, 1943.
7. C. W. Curtis, *Perforation Limits for Nondeforming Penetrators*, Frankford Arsenal Report R-903, Philadelphia, PA, 1951.
8. Frank E. Grubbs, *Linear Statistical Regression and Functional Relations*, US Army Ballistic Research Laboratories Report No. 1842, Aberdeen Proving Ground, MD, 1975.
9. D. Clark, L. Crow, and J. Sperrazza, *A Weibull Model to Estimate Residual and Critical Velocities for Target Penetrations*, US Army Materiel Systems Analysis Activity Technical Report No. 170, Aberdeen Proving Ground, MD, 1976.
10. W. Johnson, C. Collins, and F. Kindred, *A Mathematical Model for Predicting Residual Velocities of Fragments After Perforating Helmets and Body Armor*, US Army Ballistic Research Laboratories Technical Note No. 1705, Aberdeen Proving Ground, MD, 1968.
11. W. Bruchey, Jr., *A Comparison of the Residual Velocities of Various Fragment Simulating Projectiles and Actual Munitions Fragments After Penetrating Nylon Body Armor Material* (U), US Army Ballistic Research Laboratories Interim Memorandum Report No. 22, Aberdeen Proving Ground, MD, 1971 (THIS DOCUMENT IS CLASSIFIED CONFIDENTIAL.)
12. W. Kokinakis and F. H. Essig, *Penetration of Doron Body Armor Material by a Right Circular Cylinder Fragment Simulator*, US Army Ballistic Research Laboratories Memorandum Report No. 2445, Aberdeen Proving Ground, MD, 1975.
13. P. G. Morfogenis, *A Learning Curve Type Equation Predicting Residual Velocity*, US Army Ballistic Research Laboratories Memorandum Report No. 2477, Aberdeen Proving Ground, MD, 1975.
14. DARCOM-P 706-101, Engineering Design Handbook, *Army Weapon Systems Analysis, Part One*, 1977.
15. DARCOM-P 706-158, Engineering Design Handbook, *Dynamics of Ballistic Impact, Part One* (to be published).
16. J. R. Johnson and R. E. Mioduski, *Estimation of the Loss of Velocity and Mass of Fragments During Target Perforation*, US Army Ballistic Research Laboratory Memorandum Report No. 1777, Aberdeen Proving Ground, MD, 1966.
17. Joint Technical Coordinating Group for Munitions Effectiveness, 61 JTCG/ME-77-16, *Penetration Equations Handbook for Kinetic-Energy Penetrators*, November 1977. (Distributed by Tinker Air Force Base, Oklahoma City, OK. AMSAA is technical point of contact.)

CHAPTER 11

INTRODUCTION TO SELECTED TOPICS IN MULTIVARIATE STATISTICAL ANALYSIS

The analysis of random variables on which two or more characteristics are measured is introduced, and relevant topics are covered. The subject matter is approached by presenting Wilks' sample criteria and likelihood ratios for testing the equality of true means, the equality of true variances, and the equality of covariances for a multivariate normal population. An example illustrating the Wilks' theory is given for rapid firing from an M16 rifle.

Hypothesis testing based upon the analysis of results from two samples randomly selected from normal multivariate populations leads to the question concerning whether both samples originate from the same normal multivariate distribution. Therefore, a discussion is given of Hotelling's Multivariate Studentized t statistics for comparing the corresponding characteristic true means when it is assumed or known that the two samples originate from multivariate normal populations with identical variance-covariance matrices. In addition, a theoretical sketch is presented of Hotelling's Generalized T^2 statistics for comparing either the variance-covariance matrices of two normal multivariate populations or for making a simultaneous statistical judgment concerning the equality of means and the equality of variances. An example is given that compares standard artillery projectiles with an improved design.

11-0 LIST OF SYMBOLS

A_{ij} = element in the ith row and jth column of the inverse of the population variance-covariance matrix

$[A_{ij}]^{-1}$ = $[\rho_{ij}\sigma_i\sigma_j]$ = multivariate normal variance-covariance matrix

a,b,c = coefficients or constants used in Eq. 11-43 to approximate a probability level of Hotelling's T^2 statistic for any particular n, but with m taking on any value between 50 and 100

d = $(\bar{x}_1 - \bar{x}_2)$ = differences in sample means

$[d]$ = column vector of the differences in the two sample characteristic means as in Eq. 11-23

$F(\ ,\ )$ = Snedecor's "F" statistic or ratio for the number of degrees of freedom indicated before and after the comma

H_m = Wilks' hypothesis that states that the population means are all equal when it is assumed that the variances are equal and the covariances are equal

H_{mvc} = Wilks' combined or overall test of the statistical hypothesis that the true means are all equal, the variances are equal, and the covariances are equal

H_{vc} = Wilks' hypothesis that states that the population variances are equal and the population covariances are equal

$I_u(p,q)$ = Karl Pearson's incomplete beta ratio function, with argument u and parameters p and q (see Ref. 12)

i = 1, 2,..., k

k = dimension of normal multivariate or k-variate population

L_m = likelihood ratio statistic for testing Wilks' hypothesis H_m (see Eq. 11-10)

L_{mvc} = likelihood ratio statistic for testing Wilks' hypothesis H_{mvc} (see Eq. 11-12)

L_{vc} = likelihood ratio statistic for testing Wilks' hypothesis H_{vc} (see Eq. 11-11)

M = sample size for the "new" or second designated sample

m = number of degrees of freedom in the second sample

N = sample size for the "old" or first designated sample

N = total sample size

N_1 = number of items in the first sample

N_2 = number of items in the second sample

n = $(N-1)$ = number of degrees of freedom in variance estimate

n_1 = number of degrees of freedom in the first sample

n_2 = number of degrees of freedom in the second sample

p = 1, 2,..., N

r = "average" sample correlation coefficient of all k-characteristics as defined in Eq. 11-9

S^2 = estimate of variance based on the sum of the sums of squares in both samples and the total degrees of freedom (see Eq. 11-16)

$[s_{ij}]$ = denotes the variance-covariance matrix of the sample values

$|s_{ij}|$ = denotes the determinant of the variance-covariance matrix

s'_{ij} = covariance of the z_{ij}'s as defined in Eq. 11-37, but also amounts to just the covariance of the x'_{ij} or new sample values

s''_{ij} = sample covariance type quantity based on the z_{ip} not subtracted from their respective sample mean values

$\tilde{s}_{ij}$ = represents a sample covariance based on the whole sample size N as in Eq. 11-7; not the degrees of freedom n. If $j = i$, this quantity becomes a variance.

$\tilde{s}^2$ = average sample variance of the k-characteristics in Eq. 11-8

$\tilde{s}_i^2$ = $\tilde{s}_{ii}$ = sample variance based on the N sample items in Eq. 11-8

$T^2(1\%)$ = upper 1% significance level of Hotelling's Generalized T^2 statistic

T_D^2 = Hotelling's Generalized T^2 statistic for testing the equality of two variance-covariance matrices only

T_M^2 = another form of Hotelling's Multivariate Studentized t statistic and is related to T_S^2 by Eq. 11-33. Like T_S^2, T_M^2 is used to test the hypothesis that the true means of the corresponding characteristics are equal when it is assumed that the variance-covariance matrices are equal

T_S^2 = Hotelling's Multivariate Studentized t statistic for testing equality of normal multivariate population means, assuming the variance-covariance matrices are equal (see Eqs. 11-18 or 11-21 for example)

T_m^2 = value of Hotelling's Generalized T^2 statistic for m degrees of freedom. The subscript can take on values $m+1$, $m+2$, etc.

T_p^2 = pth term value of Hotelling's Generalized T^2 statistic as in Eq. 11-27

T_0^2 = Hotelling's Generalized T^2 statistic for jointly testing the equality of variance-covariance matrices and the equality of means based on two samples from multivariate normal populations

t = usual or ordinary Student's t ratio as in Eq. 11-17

tr = denotes the trace of a matrix, i.e., sum of elements of the principal diagonal of the matrix

v_{ij} = element in the ith row and jth column of $[v_{ij}]$

$[v_{ij}]$ = $[s_{ij}]^{-1}$ = inverse matrix of the sample variance-covariance matrix for a normal multivariate population

w = $T^2/(2m + T^2)$ = convenient random variable of the ratio of Hotelling Generalized T^2 statistic used in the probability distribution form of Eq. 11-41

x_{ip} = represents the pth observation of the ith characteristic of the normal multivariate sample value; sometimes shortened to x_i

x_{1p} = pth observation in the first sample

x_{2p} = pth observation in the second sample

x'_{ip} = pth sample value for the ith characteristic in the new sample
$\bar{x}$ = overall sample mean of all k-characteristics in Eq. 11-6
$\bar{x}_i$ = sample mean for the ith characteristic as in Eq. 11-6
$\bar{x}_1$ = mean of the first sample
$\bar{x}_2$ = mean of the second sample
z_{ip} = $(x'_{ip} - \bar{x}_i)$ = deviation of the pth new sample value for the ith characteristic from the sample mean of the old sample for the same or ith characteristic
$\bar{z}$ = sample mean of z's
α = small probability level
$\Gamma(\ )$ = complete gamma function of the quantity in parentheses
μ = hypothesized common value of the μ_i
μ_i = population mean of the ith characteristic x_i
μ_j = population mean of the jth characteristic x_j
μ_{i0} = a common hypothesized mean value for the characteristics of a multivariate normal population
ρ_{ij} = population correlation coefficient between x_i and x_j
Σ = sum to be taken over all sample observations
σ = hypothesized common value of the σ_i
σ_i^2 = population variance of the ith characteristic x_i
σ_j^2 = population variance of the jth characteristic x_j
σ_{ij} = $\rho_{ij}\sigma_i\sigma_j$ = population covariance of the ith and jth characteristics
$\chi^2[\ ]$ = denotes the chi-square variate for the number of degrees of freedom within the brackets

11-1 INTRODUCTION

Although the topics covered in the preceding chapters of this handbook involve the analysis of data described by and primarily following some of the key univariate probability distributions, there are a very large number of Army statistical applications that require the analysis of bivariate and multivariate or joint variables. For example, in analyses concerned with the evaluation and overall effectiveness of Army weapons or weapon systems (Refs. 1 and 2), the prime requirement is to analyze two-dimensional data, such as range and deflection errors, and occasionally the analysis of three-dimensional variations is required, such as in air defense. When one encounters the analysis of two-dimensional, or bivariate, data, such as range and deflection variations of artillery projectile impacts or the vertical and horizontal locations of rifle bullet or antitank projectile strikes, a number of parameters describing the resulting patterns arise. Moreover, it becomes important to make some comparisons of the measures of dispersion and mean locations of the distributions in the two directions. A very practical description and analysis of the patterns of shots for rifles, antitank weapons, and many missiles may be found in Ref. 3, which also contains many examples. The measures of bullet pattern tightness and location described in Ref. 3 include, for example, variances and standard deviations in the two directions, the circular error probable (CEP), the extreme horizontal and extreme vertical dispersions, the mean horizontal and mean vertical deviations, the radial standard deviation, the mean radius, the extreme spread, the radius of the covering circle, and the diagonal of the shot patterns. Analyses concerning the center of impact (C of I) locations and deviations from aim points are also illustrated. Most of these analyses are concerned, however, with "circular" patterns, i.e., the case for which the standard deviations of the fall of shots in the two directions are equal. A need exists, therefore, for the Army statistician or analyst to have at hand an account of procedures for judging the "circularity" of shot patterns and some methods of analysis for the noncircular case, especially if some dependence between the two directions exists. In fact, it is often the covariance term or "correlation" between the range and deflection errors that gives some difficulty in analyses, and there is always the need to know whether or not the data can be considered a

homogeneous sample drawn from the hypothesized bivariate normal population of prime interest. As will be seen in the sequel, some special statistical tests of significance developed by Wilks (Ref. 4) may be used to answer such questions. These tests are referred to as sample criteria for testing the equality of means, the equality of variances, and the equality of covariances in a normal (bivariate) or multivariate population.

The Wilks statistics (Ref. 4) are for either a single bivariate or multivariate normal sample. However, many important applications exist that require the analysis of data from two bivariate or multivariate samples. For example, suppose that standard production artillery projectiles exhibit certain range and deflection dispersion patterns, but it is desired to design a new artillery projectile that will give a much tighter pattern of dispersion in the two directions. It will be necessary to demonstrate that the new projectile will give an improved dispersion pattern, and one will be led to an experimental firing program, the aim of which will be to compare samples of the "old" projectile with those of the newly designed artillery projectile. Hence it becomes desirable to make inferences about range and deflection variations or to test some hypotheses concerning the relative sizes of the population variances and covariances of the "old" and the "new" projectiles. In addition, as is frequently the case, one would also like to determine whether newly designed artillery projectiles will give increased ranges—a very desirable goal indeed. Bivariate and multivariate statistical problems of this nature have considerable Army interest and have been thoroughly investigated by Hotelling (Refs. 5 through 7), Hunter (Ref. 8), Grubbs (Ref. 9), and Grubbs, Coon, Hunter, and Crowder (Ref. 10). The main stimulus for this work arose in connection with the analysis of bombing problems by Hotelling (Ref. 5) during World War II.

Although our discussions and approaches are of a military nature, applications to other activities will be readily seen.

11-2 TESTS FOR EQUALITY OF POPULATION MEANS, EQUALITY OF VARIANCES, AND EQUALITY OF COVARIANCES FOR MULTIVARIATE NORMAL DISTRIBUTIONS

For the bivariate normal population, the need exists to know whether the standard deviations in the two directions are equal, whether the true means—which determine the centroid or C of I—are equidistant from the aim point, and whether there is nonzero correlation between the variates of the two-dimensional scatter diagram. We refer to the coordinate axes as the x- and y-directions. Then if the standard deviation in x is equal to that in the y-direction, the pattern is "circular". If the pattern of shots in the firing of weapons is indeed circular, then this simplifies the problem of analysis of the data and subsequent modeling efforts. Of course, a straightforward (Snedecor-Fisher) "F" test for the observed ratio of sample variances in the two directions would ordinarily give an answer to the question of circularity. However, one could be fooled by such a test of significance if some clustering of the shots exists along a line not coincident with either of the axes. In fact, there could be quite a difference in the sigmas along an inclined axis relative to x and y, so that dependence is evident and still the projection of points onto the x- and y-axes may show equal sigmas. Thus it becomes necessary to test for dependence in the x- and y-scatter or to test for "correlation". In practical situations and for the bivariate case, this can be done usually well by a t-test of whether the population correlation coefficient is truly zero. When one also considers the problem of whether the coordinates of the C of I of the shots are located at equal distances in the two directions from the point of aim, a complete, joint test concerning the equality of means, equality of variances, and nondependence of the impact coordinates becomes very important. The Wilks tests and approach (Ref. 4) are designed to settle such questions for a k-variate or multivariate normal sample. Our prime interest will be for $k = 2$, i.e., the bivariate case. For the bivariate case note that the covariance of x and y is also the covariance of y and x, so that there is only a single covariance. However, for the k-variate population there could be several or many covariances, not all equal. Here we include results for the general k-variate case where convenient, in line with the thought that some readers will need equations for the $k \geq 2$ application. These tests are robust to hidden correlations.

Suppose we sample a normal k-variate population—for which $x_1, x_2, \ldots, x_k$ are the variates—such that μ_i is the mean of x_i, σ_i^2 is the variance of x_i, and $\rho_{ij}\sigma_i\sigma_j$ is the covariance (ρ_{ij} = population correlation coefficient) between x_i and x_j. The normal k-variate distribution law of the x_i in the population is

$$\frac{[A_{ij}]^{1/2}}{(2\pi)^{k/2}} \exp\left[-(1/2)\sum_{i,j=1}^{n} A_{ij}(x_i - \mu_i)(x_j - \mu_j)\right] \tag{11-1}$$

where the matrix of the A_{ij} is symbolized as $[A_{ij}]$, which is also symmetric, and it is the inverse of the variance-covariance matrix given by

$$[A_{ij}]^{-1} = [\rho_{ij}\sigma_i\sigma_j]\ ,\ (\rho_{ij} = 1). \tag{11-2}$$

The distribution law of Eq. 11-1 is for a "single" observation, for example, the impact point of an artillery projectile, which gives rise to both a range value and deflection position. Hence if we take the pth multivariate observation to be x_{ip}, with $i = 1, \ldots, k$ to be the dimension of the multivariate normal population, and $p = 1, \ldots, N$ to be any sample size*, then for the k-dimension distribution law of the whole sample, one would simply raise the coefficient in Eq. 11-1 to the Nth power and sum the new exponent of Eq. 11-1 over $p = 1,\ldots,N$.

The single, overall, and joint hypothesis we wish to test, in spite of any possible hidden correlations, is that the true means μ_i are all equal, i.e.,

$$\text{All } \mu_i = \mu,\ i = 1, \ldots, k \tag{11-3}$$

all the k-variances are equal, i.e.,

$$\text{All } \sigma_i^2 = \sigma^2 \tag{11-4}$$

and all the covariances are equal, i.e.,

$$\text{All } \rho_{ij}\sigma_i\sigma_j = \rho\sigma^2 \tag{11-5}$$

where the common ρ may take on values between zero and unity. Wilks (Ref. 4) has separated this composite hypothesis into three very specific hypotheses of particular interest, namely:

1. H_{mvc} = hypothesis that the means are equal, the variances are equal, and the covariances are equal
2. H_{vc} = hypothesis that the variances are equal and the covariances are equal, irrespective of the values of the means
3. H_m = hypothesis that the true means are equal when it is assumed that the variances and covariances are equal.

Wilks uses the Neyman-Pearson likelihood ratios method of testing these hypotheses, which is based on the sample statistics or values:

$$\bar{x}_i = (1/N)\sum_{p=1}^{N} x_{ip},\quad \bar{x} = (1/k)\sum_{i=1}^{k} \bar{x}_i \tag{11-6}$$

where

$\bar{x}_i$ = sample mean for the ith characteristic
$\bar{x}$ = overall sample mean of all k-characteristics

the sample covariances $\tilde{s}_{ij}$ are

$$\tilde{s}_{ij} = (1/N)\sum_{p=1}^{N}(x_{ip} - \bar{x}_i)(x_{jp} - \bar{x}_j) \tag{11-7}$$

*$n = N - 1$ is reserved for degrees of freedom (df).

the average sample variance $\tilde{s}^2$ is

$$\tilde{s}^2 = (1/k)\sum_{i=1}^{k} \tilde{s}_{ii} = (1/k)\sum_{i=1}^{k} \tilde{s}_i^2 \tag{11-8}$$

and the average sample correlation type coefficient r is defined by

$$r = \{1/[k(k-1)]\}\sum_{i \neq j=1}^{k} \tilde{s}_{ij}/s^2. \tag{11-9}$$

Finally, the sample criteria for testing H_m, H_{vc}, and H_{mvc}—the three hypotheses of interest—on the basis of likelihood ratios L_m, L_{vc}, L_{mvc} are, respectively,

$$H_m: \quad L_m = \frac{\tilde{s}^2(1-r)}{\tilde{s}^2(1-r) + \sum_{i=1}^{k}(\bar{x}_i - \bar{x})^2/(k-1)} \tag{11-10}$$

$$H_{vc}: \quad L_{vc} = \frac{|\tilde{s}_{ij}|}{(\tilde{s}^2)^k(1-r)^{k-1}[1+(k-1)r]} \tag{11-11}$$

and

$$H_{mvc}: \qquad L_{mvc} = L_{vc}(L_m^{k-1}) \tag{11-12}$$

where $|\tilde{s}_{ij}|$ is the determinant of the sample variances and the sample covariances, i.e.,

$$\left|\tilde{s}_{ij}\right| = \begin{vmatrix} \tilde{s}_{11} & \tilde{s}_{12} \\ \tilde{s}_{21} & \tilde{s}_{22} \end{vmatrix}.$$

The L sample statistics in Eqs. 11-10, -11, and -12 will range from 0 to 1, approaching 0 when the null hypothesis of each is false and approaching unity when the null hypotheses are true. Thus if any of the hypotheses, H_{mvc}, H_{vc}, or H_m, is true, the average (accidental) value of the corresponding L will be near, but less than, unity; of course, this average value would be much nearer unity than it would for the case in which the null hypothesis is false.

For the bivariate ($k = 2$) and trivariate ($k = 3$) cases, Table 11-1 gives the 5% and 1% significance levels or critical values of L_{mvc} and L_{vc}. For the overall composite hypothesis H_m, Table 11-2 gives the 5% and 1% probability levels for $k = 2, 3, 4$, and 5 dimensional cases. To reject the null hypothesis, the observed value of L must be less than the listed values.

When the sample sizes are large (perhaps greater than about $N = 30$ or 35), the L's become approximately distributed as chi-square, i.e.,

$$-N\ln L_{mvc} \approx \chi^2[(k/2)\,(k+3) - 3] \tag{11-13}$$

$$-N\ln L_{vc} \approx \chi^2[(k/2)\,(k+1) - 2] \tag{11-14}$$

and

$$-N(k-1)L_m \approx \chi^2[k-1] \tag{11-15}$$

where the quantities in the brackets of chi-square are the df.

In actual application of the test statistics, it seems reasonable to test the hypothesis H_{mvc} first, thereby determining whether the data are consistent with the overall composite hypothesis of equal means, equal variances, and equal covariances. If not, then H_{mvc} would be rejected, and the experimenter would proceed to test the hypothesis H_{vc} of equal variances and equal covariances. Then if the data are not consistent with H_{vc},

TABLE 11-1

5% AND 1% POINTS OF L_{mvc} AND L_{vc} FOR $k = 2$ AND $k = 3$ (Ref. 4)

$k = 2$					$k = 3$				
N	L_{mvc}		L_{vc}		N	L_{mvc}		L_{vc}	
	5%	1%	5%	1%		5%	1%	5%	1%
3	0.0025	0.0001	0.0062	0.0002	4	0.00029	0.00001	0.00064	0.00003
4	0.0500	0.0100	0.0975	0.0199	5	0.0095	0.0018	0.0183	0.0035
5	0.1357	0.0464	0.2285	0.0808	6	0.0358	0.0112	0.0618	0.0198
6	0.2236	0.1000	0.3416	0.1588	7	0.0736	0.0300	0.1174	0.0493
7	0.3017	0.1585	0.4307	0.2352	8	0.1165	0.0559	0.1749	0.0866
8	0.3684	0.2154	0.5005	0.3039	9	0.1603	0.0860	0.2297	0.1272
9	0.4249	0.2683	0.5559	0.3637	10	0.2028	0.1181	0.2802	0.1682
10	0.4729	0.3162	0.6007	0.4154	11	0.2432	0.1508	0.3259	0.2079
11	0.5139	0.3594	0.6375	0.4601	12	0.2808	0.1829	0.3670	0.2457
12	0.5493	0.3981	0.6682	0.4989	13	0.3157	0.2141	0.4040	0.2811
13	0.5800	0.4329	0.6943	0.5328	14	0.3480	0.2439	0.4373	0.3141
14	0.6070	0.4642	0.7165	0.5626	15	0.3778	0.2722	0.4674	0.3448
15	0.6307	0.4924	0.7358	0.5889	16	0.4052	0.2990	0.4946	0.3732
16	0.6518	0.5180	0.7528	0.6124	17	0.4306	0.3243	0.5193	0.3996
17	0.6707	0.5411	0.7675	0.6334	18	0.4540	0.3482	0.5418	0.4240
18	0.6877	0.5623	0.7807	0.6522	23	0.5484	0.4482	0.6293	0.5230
19	0.7030	0.5817	0.7925	0.6693	33	0.6660	0.5811	0.7326	0.6470
20	0.7169	0.5995	0.8031	0.6848	63	0.8135	0.7591	0.8549	0.8029
21	0.7294	0.6159	0.8126	0.6989	∞	1.0000	1.0000	1.0000	1.0000
22	0.7411	0.6310	0.8213	0.7119					
23	0.7518	0.6450	0.8292	0.7237					
24	0.7616	0.6579	0.8365	0.7347					
25	0.7707	0.6700	0.8431	0.7448					
26	0.7791	0.6813	0.8493	0.7542					
27	0.7869	0.6918	0.8549	0.7629					
28	0.7942	0.7017	0.8602	0.7710					
29	0.8010	0.7110	0.8651	0.7786					
30	0.8074	0.7197	0.8697	0.7857					
31	0.8133	0.7279	0.8739	0.7924					
32	0.8190	0.7356	0.8779	0.7987					
42	0.8609	0.7943	0.9073	0.8454					
62	0.9050	0.8577	0.9375	0.8945					
122	0.9513	0.9261	0.9684	0.9460					
∞	1.0000	1.0000	1.0000	1.0000					

Reprinted with permission. Copyright © by the Institute of Mathematical Statistics.

one finally proceeds to test the hypothesis H_m to judge whether the true means can be considered to be equal, assuming equal variances and equal covariances. This order of procedure is merely a suggestion, and most often one would like to calculate and examine all of the L's closely. In applications the main interest is usually centered on true mean values as in Student's t test.

The statistical tests of multivariate hypotheses described here will find the best applications in those cases for which the different directions or variates are in the same physical units. If, for example, the analyst is

TABLE 11-2

5% AND 1% POINTS OF L_m (Ref. 4)

$k=2$			$k=3$			$k=4$			$k=5$		
N	5%	1%	N	5%	1%	N	5%	1%	N	5%	1%
2	0.0062	0.0002	2	0.0500	0.0100	2	0.0973	0.0328	2	0.1354	0.0589
3	0.0975	0.0199	3	0.2236	0.1000	3	0.2960	0.1698	3	0.3426	0.2221
4	0.2285	0.0808	4	0.3684	0.2154	4	0.4372	0.3002	4	0.4793	0.3566
5	0.3416	0.1588	5	0.4729	0.3162	5	0.5340	0.4019	5	0.5709	0.4560
6	0.4307	0.2352	6	0.5493	0.6033	6	0.6033	0.4800	6	0.6356	0.5302
7	0.5005	0.3039	7	0.6070	0.4642	7	0.6550	0.5409	7	0.6837	0.5872
8	0.5559	0.3637	8	0.6518	0.5180	8	0.6950	0.5895	8	0.7206	0.6321
9	0.6007	0.4154	9	0.6877	0.5623	9	0.7267	0.6290	11	0.7933	0.7232
10	0.6375	0.4601	10	0.7169	0.5995	10	0.7525	0.6617	16	0.8559	0.8043
11	0.6682	0.4989	11	0.7411	0.6310	11	0.7739	0.6892	31	0.9246	0.8961
12	0.6943	0.5328	12	0.7616	0.6579	21	0.8788	0.8290	∞	1.0000	1.0000
13	0.7165	0.5626	13	0.7791	0.6813	41	0.9372	0.9101			
14	0.7358	0.5889	14	0.7942	0.7017	∞	1.0000	1.0000			
15	0.7527	0.6124	15	0.8074	0.7197						
16	0.7675	0.6334	16	0.8190	0.7356						
17	0.7807	0.6522	21	0.8609	0.7943						
18	0.7925	0.6693	31	0.9050	0.8577						
19	0.8031	0.6848	61	0.9513	0.9261						
20	0.8126	0.6989	∞	1.0000	1.0000						
21	0.8213	0.7119									
22	0.8292	0.7237									
23	0.8365	0.7347									
24	0.8431	0.7448									
25	0.8493	0.7542									
26	0.8549	0.7629									
27	0.8602	0.7710									
28	0.8651	0.7786									
29	0.8697	0.7857									
30	0.8739	0.7924									
31	0.8779	0.7987									
41	0.9073	0.8454									
61	0.9375	0.8945									
121	0.9684	0.9460									
∞	1.0000	1.0000									

Reprinted with permission. Copyright © by the Institute of Mathematical Statistics.

examining muzzle velocity (MV) and pressure data on rounds fired from the same gun, he may want to convert the pressure data into equivalent velocity data by using an appropriate physical law. For the impact positions of rounds on the ground or a vertical target, the data are already in like physical units, i.e., inches, feet, meters, etc. Wilks (Ref. 4) gives an excellent example in the field of educational psychology for which each of 100 students are examined through the use of three different tests to determine whether the test procedures constitute "parallel forms", i.e., are equally "valid". Apparently, such experiments might develop the "best" test or a standard test form. Clearly, there could well be some dependence involved because the same student takes each of the three tests, and although the test design seems very proper, this dependence requires further consideration.

For an example here we might take the pattern of artillery projectile impacts or rocket strikes on the ground for demonstrating Wilks' useful theory, but a very different type of example would be illuminating. Let our interest center on the rapid firing mode of an M16 rifle. In this connection, suppose that groups of three rounds each in rapid fire are shot from the M16. Since the first round of each group of three is well aimed—but this is hardly the case at all for the second, or even the third, bullet—does the "jump" between the first and second rounds of a group have a significant effect on the pattern tightness or accuracy? For simplicity, we will deal with only the vertical jump.

Example 11-1:

Table 11-3 lists the vertical points of impact x_1 and x_2 in centimeters for the first bullet and the second bullet, respectively, for 10 three-round rapid groups fired from the M16 rifle. Since the strike of the second round may be well correlated with the impact of the first aimed bullet, what can be said about possible changes in the average vertical impact and the variability characteristics of the first two bullets of a group? Also is there any evidence that the point of impact of the second round depends highly on that of the first bullet strike, or is the infantryman able to re-aim the M16 between rounds?

TABLE 11-3

VERTICAL DEVIATIONS FROM AIM POINT OF FIRST AND SECOND BULLETS FIRED IN RAPID-FIRE GROUPS OF THREE ROUNDS WITH M16 RIFLE*

First Bullet Impact x_1, cm	Second Bullet Impact x_2, cm
–1.59	–2.32
–0.17	3.96
–1.84	3.77
–0.98	4.39
–0.62	6.73
–0.32	–0.06
–0.98	1.66
–1.30	–0.31
–0.39	4.15
–1.25	0.92

For significance tests of the hypotheses—H_{mvc}, H_{vc}, and H_m—the sample statistics of calculable interest are

$\bar{x}_1 = -0.944$ $\quad \bar{x}_2 = 2.289$ $\quad \bar{x} = 0.6725$ (by Eq. 11-6)

$\tilde{s}_1^2 = 0.2843$ $\quad \tilde{s}_2^2 = 6.8371$ $\quad \tilde{s}^2 = 3.5607$ (by Eqs. 11-7 and 11-8)

$\tilde{s}_{12} = 0.5239$ (by Eq. 11-7) $\quad r = 0.1471$ (by Eq. 11-9) $\quad |\tilde{s}_{ij}| = 1.6693$ (by Eq. 11-11)

Then it is found that

$L_{mvc} = 0.176$ $\quad L_{vc} = 0.479$ $\quad L_m = 0.368.$

Observe the 5% upper probability levels of Table 11-1 for $k = 2$ for L_{mvc} and L_{vc} and also those in Table 11-2 for L_m with $N = 10$. Note that all of the observed L's are less than the corresponding tabular values. In fact, significance is established even with respect to the 1% probability levels for L_m and L_{mvc}. Thus this particular sample of bivariate data does not support any of the three hypotheses. Therefore, they are rejected with the

*We are pleased to acknowledge the suggestion to use these data furnished by Mr. Weldon Willoughby and Mr. Robert Eissner of the US Army Materiel Systems Analysis Activity (AMSAA).

conclusion that the mean points of impact of the first and second rounds are quite different, as are the variances. For the bivariate case there is only a single covariance; therefore, the overall test does not really check on any comparison. Nevertheless, the two parts of the bivariate normal population are not the same; the mean impact of the second bullet jumps $2.29 + 0.94 = 3.23$ cm above that of the first bullet with an inclination to the upper right, and the ratio of sigmas is estimated to be $(6.84/0.28)^{1/2} = 4.94$.

Although we have analyzed only the data for the first two rounds and only the vertical direction in this example, the reader will note that since each group consisted of three rapidly fired bullets, there is a complete trivariate normal sample with three coordinate impacts. (The spatial trivariate case collapses to coplanar impacts on an xy- plane.) Hence an easy extension of our analysis to the three-bullet target firings could be carried out by using the Wilks theory (Ref. 4) covered in this paragraph.

The Wilks theory of Ref. 4 is especially valuable for making statistical judgments concerning the "circularity" of bivariate distributions or the "sphericity" of multivariate distributions and for comparing the location parameters or true means of the component distributions. Again, however, we note that the hypotheses tested are for single multivariate samples. The next logical step, therefore, would involve the comparison of two multivariate samples—which it is somewhat natural to refer to as the "old" and the "new" samples—in order to detect any change or shift in the population parameters. For the two-sample cases we are led to consider the use of Hotelling's Multivariate Studentized t-type statistic (Ref. 6)* and Hotelling's Generalized T^2 measures of multivariate dispersion. These tests are discussed in par. 11-3.

11-3 SELECTED TOPICS AND APPLICATIONS OF HOTELLING'S MULTIVARIATE STUDENTIZED t RATIOS AND GENERALIZED T^2 STATISTICS

In our presentation it is desirable to cover the Hotelling Generalized Student's t ratios and the generalized T^2 measures of multivariate dispersion in separate subparagraphs.

11-3.1 HOTELLING'S GENERALIZATION OF THE STUDENT-FISHER t RATIOS

In Chapter 4 we gave a suitably complete account of the Student's t tests for univariate samples from a normal population. One of the significance tests was based on a single normal sample, and we used the Student's t ratio of the difference between the observed sample mean and a hypothesized population mean divided by the estimated standard deviation of the difference to judge the true location of the assumed normal population. The other case involved two samples and tested the hypothesis that both samples were taken from the same normal population once it had been established that the variances were equal. This particular Studentized t ratio consisted of the difference of the two sample means divided by the estimated standard deviation of that difference. In case the two variances were judged to be different, one might still be interested in judging whether the two normal population means are coincident, which involves the Behrens-Fisher ratio test (see par. 4-7.3.2).

A natural, instructive approach toward the uses of Hotelling's Generalized Student's t ratio, or multivariate T^2 as it is called, is to start with the two-sample, univariate case and then to generalize that statistic to the k-variate, or multivariate, case. This means that we amplify the Student's t statistic of Eq. 4-108. For the purposes of this chapter and the consistency of notation therein, we define the following:

N_1 = number of observations in the first sample
n_1 = $(N_1 - 1)$ = df in first sample (In par. 11-3.2, n will be used for the bivariate case.)
N_2 = number of observations in the second sample
n_2 = $(N_2 - 1)$ = df in second sample (In par. 11-3.2, m will be used for the bivariate case.)
x_{1p} = pth observation in the 1st sample, $p = 1,\ldots,N_1$
x_{2p} = pth observation in the second sample, $p = 1,\ldots,N_2$
$\bar{x}_1$ = mean of the first sample
$\bar{x}_2$ = mean of the second sample

*There is also a special test of Hotelling's Multivariate Studentized t for a single multivariate sample as discussed in par. 11-3.1.

d = $\bar{x}_1 - \bar{x}_2$ = difference in sample means
Σ = sum over entire sample, i.e., 1 to N_1 or 1 to N_2
S^2 = estimated sample variance based on both samples and total number of df

$$= \frac{\Sigma[x_{1p} - \bar{x}_1)^2 + \Sigma(x_{2p} - \bar{x}_2)^2}{N_1 + N_2 - 2} \tag{11-16}$$

where $p = 1$ to N_1 for the first summation, and $p = 1$ to N_2 for the second summation. The ordinary Student's t ratio for testing the equality of two normal population means then is given by

$$t = d/[S(1/N_1 + 1/N_2)^{1/2}] \tag{11-17}$$

and this may be rewritten in the form of Hotelling's Multivariate Studentized ratio as

$$T_S^2 = t^2 = \left(\frac{N_1 N_2}{N_1 + N_2}\right)[d]^T [S^2]^{-1}[d]$$
$$= \left(\frac{N_1 N_2}{N_1 + N_2}\right) d(S^2)^{-1} d\,. \tag{11-18}$$

By comparing Eqs. 11-17 and 11-18, we might say that Eq. 11-17 is in a "linear" form, whereas Eq. 11-18 is in a "square" form. That is to say, whereas Eq. 11-17 is distributed in probability as Student's t with $(N_1 + N_2 - 2)$ df, the square of t, or Eq. 11-18, follows the Snedecor F, or variance ratio, distribution with 2 and $(N_1 + N_2 - 2)$ df. The quantity T_S^2 is known as Hotelling's Studentized t statistic for the bivariate case although we have applied it to two univariate samples rather than to the two different orthogonal directions for the bivariate sample case. Nevertheless, the form of Eq. 11-18 generalizes to the Hotelling Multivariate Studentized statistic, and we have used the subscript "S" to distinguish it from Hotelling's Generalized T^2 that is used to compare the dispersion matrices (variance-covariance matrices) of two normal multivariate samples. An example seems in order.

Example 11-2:

With reference to the data of Example 11-1, a large disparity was noticed in mean points of impact for the first and second bullets, and the variances of the two bullets were widely different. Despite the different standard errors of the two bullets (and the correlation between the impacts of the two bullets), can the Hotelling Studentized t statistic of Eq. 11-18 detect the difference in mean impact points if one treats the data as two univariate samples?

For this example we have $N_1 = N_2 = 10$, $d = \bar{x}_1 - \bar{x}_2 = -3.233$ and from Eq. 11-16

$$S^2 = [9(0.2843) + 9(6.8371)]/18 = 3.561.$$

Then from Eq. 11-18 we obtain

$$T_S^2 = t^2 = 14.68$$

but the upper 5% F probability level for 1 and 18 df is only 4.41, so there is indeed a great jump upward for the second bullet. Moreover, we are impressed by the robustness of the test.*

*Nevertheless, the Behrens-Fisher test of par. 4-7.3.2 would be more appropriate here, but it would still render high significance.

Having generalized the ordinary univariate Student's t statistic to its analogue form (Eq. 11-18) for Hotelling's Multivariate Student's T_S^2, we are in a position to discuss some other properties for either the bivariate or multivariate normal populations. Let us first consider the case of a single bivariate or multivariate sample. Here we take the quantities or observations $x_{1p}, x_{2p}, \ldots, x_{kp}$ to represent the pth sample item from a normally correlated multivariate population, with $p = 1, \ldots, N$ random sample elements. Thus the pth sample value has k mutually related characteristics—e.g., height, weight, and arm length—of humans. Suppose further that the true or hypothesized means of the k-characteristics are $\mu_1, \mu_2, \ldots, \mu_k$, and we take $[s_{ij}]$ to be the matrix of unbiased estimates of the true covariances σ_{ij} (and variances $\sigma_{ii} = \sigma_i^2$), where for $n = (N - 1)$ df we have

$$s_{ij} = (1/n) \sum_{p=1}^{N} (x_{ip} - \bar{x}_i)(x_{jp} - \bar{x}_j). \tag{11-19}$$

Finally, let us define the inverse of the sample variance-covariance matrix to be

$$[v_{ij}] = [s_{ij}]^{-1}. \tag{11-20}$$

Then the quantity or quadratic form given by

$$T_S^2 = \sum_{1}^{k} \sum_{1}^{k} v_{ij} (\bar{x}_i - \mu_i)(\bar{x}_j - \mu_j) N \tag{11-21}$$

is distributed in probability as Hotelling's Multivariate Student's T^2 statistic, and the transformed statistic

$$F(k, n - k + 1) = (n - k + 1)\ T_S^2/(kn) \tag{11-22}$$

is distributed as Snedecor's F or variance ratio with k and $(n - k + 1)$ df. Hence one could use Eq. 11-21 for a single multivariate normal sample to test the hypothesis that the true unknown component means of the mutual characteristics take on specified values μ_{i0}, either all equal values or different values.

A very important and useful application of Hotelling's Multivariate Studentized t statistic is in connection with two normal samples for either the bivariate or the multivariate case. The object is to compare the corresponding true means of the different characteristics of the normally correlated samples when it is assumed that the two samples originate from two multivariate normal populations having identical variance-covariance matrices. This is easily done with a very straightforward extension of the first right-hand side (RHS) of Eq. 11-18. It is convenient in this chapter to call the first of the two samples the "old" sample, which consists of N items or observations (or $n = (N - 1)$ df). To distinguish the second sample, we could list the observations as x'_{ip}, i.e., the same notation except that "primes" are used, and we call the second sample the "new" sample, which has M observations or $m = (M - 1)$ df for the estimated variance. Moreover, since it is assumed that the two normally correlated samples originate from populations at least with identical variance-covariance matrices, then for the estimated variances and covariances based on sample values we may "pool" the sums of squares (SS)—or cross products—similar to those of Eq. 11-16 and divide by the correct number of df, or $(n + m) = (N + M - 2)$. Finally, the column vector defining the difference d of Eq. 11-18 becomes

$$[d] = \begin{bmatrix} \bar{x}_1 - \bar{x}'_1 \\ \bar{x}_2 - \bar{x}'_2 \\ \vdots \\ \bar{x}_k - \bar{x}'_k \end{bmatrix} \tag{11-23}$$

Hotelling's Multivariate Studentized t or T_S^2 then becomes

$$T_S^2 = \frac{NM}{N + M} [d]^T [s_{ij}]^{-1} [d] = \frac{NM}{N + M} [d]^T [v_{ij}] [d] \tag{11-24}$$

where the variance-covariance matrix $[s_{ij}]$ has $(n+m)$ df, and $[v_{ij}]$ is its inverse. In referring an observed value of T_S^2 to a table of significance levels, one uses

$$F(k, n+m-k+1) = \left(\frac{m+n-k+1}{mk+nk}\right) T_S^2 \tag{11-25}$$

which is distributed in probability as Snedecor's F with k and $(m+n-k+1)$ df. We will illustrate the use of Hotelling's Multivariate T_S^2 in Example 11-3 for a two-sample bivariate case requiring also the application of Hotelling's Generalized T^2 statistics. In fact, it is best to apply Hotelling's Multivariate T_S^2 statistic for mean values after we have established that two samples drawn from normal multivariate populations have equivalent variances and covariances, respectively.

11-3.2 HOTELLING'S GENERALIZED T^2 STATISTICS

The main thrust of the analysis concerning Hotelling's Generalized T^2 statistics is that of determining whether or not two normal multivariate samples originate from the same multivariate normal population, i.e., whether the corresponding true means are equal and their variances and covariances are the same, respectively. We will approach this problem primarily by illustrating the bivariate case although it will be quite clear that an extension to any number of dimensions k is very obvious. As an example, we will make a comparison of the range and deflection patterns of a standard "old" type of artillery projectile and a proposed, or "new", artillery projectile to replace the "old" one.

Approaching the Hotelling Generalized T^2 statistics from the bivariate form, we start with the old sample of N items, with means of the characteristics equal to $\bar{x}_1$ and $\bar{x}_2$, and the sample variance-covariance matrix of the old sample $[s_{ij}]$, which is based on $n = (N-1)$ df. We then label the M new sample values as x'_{ip}, $i = 1, 2$ (or i running from 1 to k) and $p = 1, 2, \ldots, M$. New deviations or z's are generated, which are determined from

$$z_{ip} = x'_{ip} - \bar{x}_i \tag{11-26}$$

which, for the bivariate case (or i from 1 to k), give M residuals of the new sample values from the old sample means. The Hotelling Generalized T^2 statistic for testing the conformance of the pth new sample value to the population of the old sample values is then

$$T_p^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} v_{ij} z_{ip} z_{jp}, \quad (k = 2 \text{ here}) \tag{11-27}$$

where we have that v_{ij} are the elements of the inverse variance-covariance matrix of the old sample, $[s_{ij}]^{-1}$. It is of interest to note that the quantities z_{ip} not only contain the individual residuals and average out to the difference in means of new and old sample values, but also actually contain relevant information on dispersion of the new sample since the old sample means amount to constants anyway (in calculating the variance of the z_{ip} with respect to p). Hence the total characterization of the conformance of the entire new sample to the old bivariate or multivariate normal sample will be given by

$$T_0^2 = T_1^2 + T_2^2 + \cdots + T_p^2 + \cdots + T_M^2. \tag{11-28}$$

Whereas one notes in particular that Eq. 11-28 adds a generalized T^2 for each and every sample point, Hotelling in Refs. 5 and 7 divides the total T_0^2 into two more pertinent parts or quantities. These two parts are more useful in comparing the variance-covariance matrices of the two samples with one T^2 statistic, followed by a direct comparison of mean values with the other Hotelling statistic, which incidentally is a Hotelling Multivariate Studentized statistic. This division of T_0^2 into two parts is based on the identity

$$\sum_{p=1}^{M} z_{ip} z_{jp} = \sum_{p=1}^{M} (z_{ip} - \bar{z}_i)(z_{jp} - \bar{z}_j) + M \bar{z}_i \bar{z}_j. \tag{11-29}$$

With the use of Eq. 11-29, it is found that the quantity T_0^2 may be expressed as

$$T_0^2 = T_D^2 + T_M^2 \tag{11-30}$$

with

$$T_D^2 = \sum_{i=1}^{k=2} \sum_{j=1}^{k=2} v_{ij} \sum_{p=1}^{M} (z_{ip} - \bar{z}_i)(z_{jp} - \bar{z}_j) \tag{11-31}$$

and

$$T_M^2 = M \sum_{i=1}^{k=2} \sum_{j=1}^{k=2} v_{ij} \bar{z}_i \bar{z}_j \tag{11-32}$$

where for the bivariate case of our prime interest $k = 2$.

Special attention should be given to the upper limits in Eqs. 11-31 and 11-32. We have terminated the summations at $k = 2$, or the bivariate case, since our main interest is the two-dimensional case and available exact distribution theory more or less ends for $k = 2$.

Some particular emphasis should be placed on the fact that in Eq. 11-31 each of the $(z_{ip} - \bar{z}_i)$ may be replaced by $(x'_{ip} - \bar{x}'_i)$, the residuals or deviations from the sample mean of the new sample observations. Hence the quantity T_D^2 in Eq. 11-31 actually represents a comparison between the covariances of the new bivariate normal sample or x' values and the old sample values x since the v_{ij} are the elements in the inverse matrix of the old sample variance-covariance matrix. This quantity T_D^2 follows Hotelling's Generalized T^2 probability distribution for the bivariate case as is indicated in Eq. 11-41 that follows, and the upper 5% and 1% probability levels are given in Ref. 10 along with an approximation. The T_D^2 and T_M^2 so computed can be added to give T_0^2, whereas a check in the computations may be obtained by using z_{ip} and calculating T_0^2 directly as in Eq. 11-28. The T_D and T_M are not independent since they depend on the same old sample; however, their conditional distributions are independent for a particular old sample as shown by Hotelling (Ref. 7).

The quantity T_M^2 of Eq. 11-32, which must be used in terms of the z's only (as in Eq. 11-26), follows Hotelling's Multivariate Student T_S^2 distribution as introduced in par. 11-3.1. In fact, for the bivariate case $k = 2$, the relation between T_S^2 and T_M^2 is given by

$$T_S^2 = NT_M^2/(N + M) \tag{11-33}$$

and we may use the Snedecor F variate

$$F(2, N-2) = N(N-2)T_M^2/[2(N+M)(N-1)]^* \tag{11-34}$$

which is distributed as F with 2 and $(N-2)$ df. Hence we have available a relatively simple significance test for the quantity T_M^2. For the general k-variate case, we have—using only the old sample observations in the s_{ij}—that

$$F(k, N-k) = N(n-k)T_M^2/[k(N+M)(N-1)] \tag{11-35}$$

follows the Snedecor F distribution with k and $(n - k + 1)$ df.

A further, pertinent remark concerns the large-sample or population values of the s_{ij} and, hence, v_{ij}. If the variance-covariance matrix of the old population sampled is accurately known, i.e., one has a very stable value of $[\sigma_{ij}]$, the T^2's of Eq. 11-30 become chi-squares, and in fact we have

$$\chi_0^2(2M) = \chi_D^2(2M-2) + \chi_M^2(2) \tag{11-36}$$

*If both sample SS are used to obtain $[v_{ij}]$, use $F(2, N+M-3) = N(N+M-3)T_M^2/[2(N+M)(N+M-2)]$.

where the chi-squares on the RHS are independent, and the total of $2M$ df is split into $(2M-2)$ and 2 df. This information is of value for the case where some extensive experience is available from previous testing.

Finally, we record some of Hotelling's theory for the T_0^2 and T_D^2 statistics since in applications we need to know their percentage points. Hotelling shows that for the three new sample statistics defined from

$$s'_{ij} = \left(\frac{1}{M-1}\right)\sum_{p=1}^{M}(z_{ip}-\bar{z}_i)(z_{jp}-\bar{z}_j),\ i,j=1,2 \tag{11-37}$$

i.e., two sample variances and one covariance of the new sample, these quantities have the joint Wishart (Ref. 11) probability distribution with $(M-1)$ df. For a new sample drawn independently from the same bivariate normal population as the old sample, the distribution of T_D^2 has exactly the same form as that of T_0^2 with the total new sample size M replaced by $(M-1)$ df. Therefore, for the distribution of either the quantity T_D^2 or the total T_0^2, one is interested in the distribution of Hotelling's Generalized T^2 statistic of the general form

$$T^2 = m\sum_{i=1}^{2}\sum_{j=1}^{2} v_{ij}s'_{ij} \tag{11-38}$$

where m is a general number of df for the variance-covariance variates s'_{ij}, which have the Wishart distribution (Ref. 11). Recall that n is the number of df for the old sample. An important and major result of Hotelling is that the trace, i.e., the sum of the principal diagonal elements, of the product matrix given by

$$[v_{ij}]\,[s'_{ij}] \tag{11-39}$$

is equal to T^2/m, or Hotelling's Generalized T^2 divided by m df. Hence with the use of Eq. 11-39, we are able to conduct a significance test or hypothesis test for T_D^2 that compares the relative sizes of the variance-covariance matrices of the old and new samples, and we can also carry out a significance test on T_0^2, the total dispersion matrix value, including comparisons of means.

With regard to probability distribution theory and percentage points of the generalized T^2 statistics, Hotelling (Ref. 7) uses the quantity

$$w = T^2/(2m+T^2) \tag{11-40}$$

and shows that the probability α of w being exceeded is

$$\alpha = 1 - I_w(m-1,n) + \sqrt{\pi}\left[\frac{\Gamma\left(\frac{m+n-1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\right]\left(\frac{1-w}{1+w}\right)^{(n-1)/2} I_{w^2}\left(\frac{m-1}{2},\frac{n-1}{2}\right) \tag{11-41}$$

where

$I_w(\ ,\ )$ = Karl Pearson's incomplete beta function ratio (Ref. 12)
$\Gamma(\ )$ = complete gamma function of the quantity in parentheses.

Extensive tables of the 1% and 5% probability levels of Hotelling's Generalized T^2 were originally developed at the US Army Ballistic Research Laboratories (BRL) in 1954 (Ref. 9); however, it was discovered that for values of m much greater than n some computational errors occurred in the computations due to a somewhat inaccurate approximation to the incomplete beta function ratio. Upon discovering this computational error, new 1% and 5% points were calculated for T^2, and accurate values were given in Ref. 10. The percentage points calculated are for the bivariate case, $k = 2$, only. Values of m and n, the df for the covariances of new and old samples, respectively, range over $m, n = 1(1)100$. It is not practical to include these very extensive tables in this handbook, particularly since suitable approximations can be given. In his original study of the Hotelling Generalized T^2 statistics at BRL during the summer of 1952, Prof. J. Stuart Hunter (Ref. 8) noticed that for fixed n the 5% probability levels of T^2 were practically linear with the parameter m.

Such an occurrence was hardly expected at all! However, Prof. Hunter also discovered that the linear relation was very well-established for the 1% probability levels. This fortuitous occurrence was later investigated by Helen J. Coon, formerly of the BRL, and is in a "Comment" in Ref. 10. Coon established the linear relationship in the T^2 by showing that for all m the following equation holds:

$$T^2_{m+1} - T^2_m = T^2_{m+2} - T^2_{m+1} = 2(n + T^2_m)/(2m + n - 3). \tag{11-42}$$

Therefore, since the first differences are constant, the approximate 100α percentage points, or α probability levels, of Hotelling's Generalized T^2 for a fixed n are linearly related.

The complete set of tables in Ref. 10 for m, $n = 1(1)100$ consists of about 64 pages and is accurate to practically five significant figures, as discussed by H. K. Crowder in a computational Appendix (Ref. 10). Prof. Hunter also gives two nomographs, which may be used to read off either the 1% or 5% critical values of T^2 in another Appendix of Ref. 10. For values of m greater than 50 and up to 100 (the extent of the computations), Hotelling's T^2 percentage points can be determined very accurately for fixed n from a quadratic in m. As would be expected in view of the linearity relation, the coefficient of the square term in m is quite small. The quadratic equation is

$$T^2 \approx am^2 + bm + c, \quad 50 < m \leq 100 \tag{11-43}$$

so that the appropriate set of coefficients—a, b, and c—for each n from 2 to 100 can be used to obtain accurate 1% and 5% probability levels for m greater than 50. This reduces the necessary size of the tables drastically, especially since only four pages are required for this region.

Questions of the compactness of a table and the number of significant figures to list always arise. Also it is not known just what compromises should result from the many probable applications of Hotelling's T^2 statistics. However, for Example 11-3 it would seem that three significant figures, and certainly four, should suffice. For our tabulation of the percentage points, we have decided to include 14 pages to cover the 1% and 5% points to five significant figures for m and n ranging over 1 to 50, and four pages to list values of the coefficients needed for m and n from 51 to 100, but we also list coefficients for n less than 51. Thus Table 11-4 gives the T^2 percentage points for m, $n = 1(1)50$; Table 11-5 contains the value of the coefficients a, b, and c recommended for values of m exceeding 50. These two tables should suffice.

If one is interested in the significance of the quantity T^2_D, he may calculate it using Eq. 11-38 or the trace of Eq. 11-39 and enter Table 11-4 with $n = (N - 1)$ and $m = (M - 1)$ to determine whether the observed T^2_D exceeds the tabular value. On the other hand, for the total T^2 or T^2_0—which is a combined test of whether the variance-covariance matrices are equal *and* the corresponding true means are also equal, i.e., whether the two bivariate samples are from the same normal bivariate population—we may calculate T^2_0 from Eq. 11-30 and compare the resulting value with the tabular one using $n = N - 1$ but taking $m = M$, the new sample size. Alternatively for T^2_0, we could define the covariance-like quantity s''_{ij}

$$s''_{ij} = (1/M) \sum_{p=1}^{M} z_{ip} z_{jp} \tag{11-44}$$

and calculate T^2_0 from

$$T^2_0 = M \sum_{i=1}^{2} \sum_{j=1}^{2} v_{ij} s''_{ij} = M\mathrm{tr}\{[v_{ij}][s''_{ij}]\}. \tag{11-45}$$

It is seen, in view of this discussion, that Hotelling's theory is quite complete in dealing with bivariate and multivariate statistical problems of wide interest. Since there are many facets of the overall statistical analysis and a variety of hypothesis-testing procedures, we have selected an example that should be quite informative in illustrating the Hotelling Generalized T^2 and Multivariate Studentized t statistical theory. A primary purpose is to compare the range and deflection patterns for ground impacts of some standard and proposed artillery projectiles.

TABLE 11-4

UPPER 1% AND 5% PROBABILITY OR SIGNIFICANCE LEVELS FOR HOTELLING'S GENERALIZED T^2 STATISTICS (Bivariate Case)

1% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

m \ n	2	3	4	5	6	7	8
2	0.49346 5	0.59697 3*	0.14855 3	0.76589 2	0.52237 2	0.40776 2	0.34304 2
3	0.79998 5	0.89695 3	0.21308 3	0.10670 3	0.71382 2	0.54978 2	0.45803 2
4	0.11103 6	0.11969 4	0.27698 3	0.13628 3	0.90069 2	0.68766 2	0.56917 2
5	0.14222 6	0.14969 4	0.34058 3	0.16561 3	0.10853 3	0.82345 2	0.67834 2
6	0.17349 6	0.17969 4	0.40402 3	0.19479 3	0.12686 3	0.95801 2	0.78633 2
7	0.20480 6	0.20969 4	0.46737 3	0.22388 3	0.14511 3	0.10918 3	0.89357 2
8	0.23614 6	0.23968 4	0.53066 3	0.25292 3	0.16330 3	0.12251 3	0.10003 3
9	0.26749 6	0.26968 4	0.59390 3	0.28191 3	0.18146 3	0.13579 3	0.11066 3
10	0.29886 6	0.29968 4	0.65712 3	0.31088 3	0.19958 3	0.14905 3	0.12127 3
11	0.33024 6	0.32968 4	0.72031 3	0.33982 3	0.21769 3	0.16229 3	0.13185 3
12	0.36162 6	0.35968 4	0.78348 3	0.36875 3	0.23578 3	0.17552 3	0.14242 3
13	0.39301 6	0.38967 4	0.84664 3	0.39767 3	0.25386 3	0.18872 3	0.15297 3
14	0.42440 6	0.41967 4	0.90979 3	0.42657 3	0.27192 3	0.20192 3	0.16352 3
15	0.45580 6	0.44967 4	0.97293 3	0.45546 3	0.28998 3	0.21511 3	0.17405 3
16	0.48720 6	0.47967 4	0.10361 4	0.48435 3	0.30803 3	0.22830 3	0.18458 3
17	0.51860 6	0.50967 4	0.10992 3	0.51323 3	0.32607 3	0.24147 3	0.19510 3
18	0.55000 6	0.53966 4	0.11623 4	0.54211 3	0.34411 3	0.25464 3	0.20561 3
19	0.58140 6	0.56966 4	0.12254 4	0.57098 3	0.36215 3	0.26781 3	0.21612 3
20	0.61281 6	0.59966 4	0.12885 4	0.59985 3	0.38018 3	0.28097 3	0.22663 3
21	0.64422 6	0.62966 4	0.13517 4	0.62872 3	0.39820 3	0.29413 3	0.23713 3
22	0.67562 6	0.65966 4	0.14148 4	0.65758 3	0.41623 3	0.30729 3	0.24763 3
23	0.70703 6	0.68965 4	0.14779 4	0.68645 3	0.43425 3	0.32044 3	0.25813 3
24	0.73844 6	0.71965 4	0.15410 4	0.71530 3	0.45227 3	0.33360 3	0.26862 3
25	0.76985 6	0.74965 4	0.16041 4	0.74416 3	0.47029 3	0.34675 3	0.27911 3
26	0.80126 6	0.77965 4	0.16672 4	0.77302 3	0.48831 3	0.35989 3	0.28961 3
27	0.83267 6	0.80965 4	0.17303 4	0.80187 3	0.50632 3	0.37304 3	0.30009 3
28	0.86408 6	0.83964 4	0.17934 4	0.83072 3	0.52434 3	0.38619 3	0.31058 3
29	0.89549 6	0.86964 4	0.18565 4	0.85958 3	0.54235 3	0.39933 3	0.32107 3
30	0.92690 6	0.89964 4	0.19196 4	0.88843 3	0.56036 3	0.41247 3	0.33155 3
31	0.95831 6	0.92964 4	0.19827 4	0.91728 3	0.57837 3	0.42561 3	0.34204 3
32	0.98972 6	0.95964 4	0.20458 4	0.94612 3	0.59638 3	0.43876 3	0.35252 3
33	0.10211 7	0.98963 4	0.21089 4	0.97497 3	0.61439 3	0.45190 3	0.36300 3
34	0.10525 7	0.10196 5	0.21720 4	0.10038 4	0.63240 3	0.46503 3	0.37348 3
35	0.10840 7	0.10496 5	0.22350 4	0.10327 4	0.65041 3	0.47817 3	0.38396 3
36	0.11154 7	0.10796 5	0.22981 4	0.10615 4	0.66841 3	0.49131 3	0.39444 3
37	0.11468 7	0.11096 5	0.23612 4	0.10904 4	0.68642 3	0.50445 3	0.40492 3
38	0.11782 7	0.11396 5	0.24243 4	0.11192 4	0.70443 3	0.51758 3	0.41540 3
39	0.12096 7	0.11696 5	0.24874 4	0.11480 4	0.72243 3	0.53072 3	0.42588 3
40	0.12410 7	0.11996 5	0.25505 4	0.11769 4	0.74044 3	0.54386 3	0.43635 3
41	0.12724 7	0.12296 5	0.26136 4	0.12057 4	0.75844 3	0.55699 3	0.44683 3
42	0.13039 7	0.12596 5	0.26767 4	0.12346 4	0.77644 3	0.57013 3	0.45731 3
43	0.13353 7	0.12896 5	0.27398 4	0.12634 4	0.79445 3	0.58326 3	0.46778 3
44	0.13667 7	0.13196 5	0.28029 4	0.12923 4	0.81245 3	0.59639 3	0.47826 3
45	0.13981 7	0.13496 5	0.28659 4	0.13211 4	0.83045 3	0.60953 3	0.48873 3
46	0.14295 7	0.13796 5	0.29290 4	0.13499 4	0.84846 3	0.62266 3	0.49921 3
47	0.14609 7	0.14096 5	0.29921 4	0.13788 4	0.86646 3	0.63579 3	0.50968 3
48	0.14923 7	0.14396 5	0.30552 4	0.14076 4	0.88446 3	0.64892 3	0.52015 3
49	0.15238 7	0.14696 5	0.31183 4	0.14365 4	0.90246 3	0.66206 3	0.53063 3
50	0.15552 7	0.14996 5	0.31814 4	0.14653 4	0.92046 3	0.67519 3	0.54110 3

*A tabulated value such as 0.59697 3 means 0.59697×10^3 or 596.97.

n = df for old sample
m = df for new sample

(cont'd on next page)

TABLE 11-4 (cont'd)

1% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

n/m	9	10	11	12	13	14	15
2	0.30206 2	0.27401 2	0.25372 2	0.23840 2	0.22645 2	0.21689 2	0.20907 2
3	0.40037 2	0.36114 2	0.33289 2	0.31165 2	0.29515 2	0.28197 2	0.27122 2
4	0.49504 2	0.44478 2	0.40869 2	0.38162 2	0.36063 2	0.34390 2	0.33027 2
5	0.58780 2	0.52657 2	0.48267 2	0.44981 2	0.42435 2	0.40408 2	0.38759 2
6	0.67943 2	0.60723 2	0.55555 2	0.51689 2	0.48697 2	0.46318 2	0.44383 2
7	0.77030 2	0.68715 2	0.62768 2	0.58323 2	0.54885 2	0.52153 2	0.49932 2
8	0.86066 2	0.76655 2	0.69930 2	0.64905 2	0.61021 2	0.57935 2	0.55428 2
9	0.95063 2	0.84558 2	0.77052 2	0.71448 2	0.67117 2	0.63677 2	0.60883 2
10	0.10403 3	0.92431 2	0.84146 2	0.77962 2	0.73183 2	0.69389 2	0.66307 2
11	0.11298 3	0.10028 3	0.91217 2	0.84452 2	0.79226 2	0.75076 2	0.71707 2
12	0.12191 3	0.10812 3	0.98270 2	0.90923 2	0.85249 2	0.80744 2	0.77087 2
13	0.13083 3	0.11593 3	0.10531 3	0.97380 2	0.91257 2	0.86397 2	0.82451 2
14	0.13973 3	0.12374 3	0.11233 3	0.10382 3	0.97253 2	0.92036 2	0.87801 2
15	0.14863 3	0.13154 3	0.11935 3	0.11026 3	0.10324 3	0.97664 2	0.93139 2
16	0.15752 3	0.13933 3	0.12636 3	0.11668 3	0.10921 3	0.10328 2	0.98468 2
17	0.16640 3	0.14711 3	0.13336 3	0.12310 3	0.11518 3	0.10889 3	0.10379 3
18	0.17528 3	0.15489 3	0.14035 3	0.12951 3	0.12114 3	0.11450 3	0.10910 3
19	0.18415 3	0.16266 3	0.14734 3	0.13592 3	0.12710 3	0.12009 3	0.11441 3
20	0.19301 3	0.17043 3	0.15433 3	0.14232 3	0.13305 3	0.12569 3	0.11971 3
21	0.20188 3	0.17819 3	0.16131 3	0.14871 3	0.13899 3	0.13127 3	0.12501 3
22	0.21074 3	0.18595 3	0.16828 3	0.15511 3	0.14493 3	0.13686 3	0.13030 3
23	0.21960 3	0.19371 3	0.17526 3	0.16150 3	0.15087 3	0.14244 3	0.13559 3
24	0.22845 3	0.20147 3	0.18223 3	0.16788 3	0.15681 3	0.14801 3	0.14088 3
25	0.23730 3	0.20922 3	0.18920 3	0.17427 3	0.16274 3	0.15359 3	0.14616 3
26	0.24615 3	0.21697 3	0.19616 3	0.18065 3	0.16867 3	0.15916 3	0.15144 3
27	0.25500 3	0.22472 3	0.20313 3	0.18703 3	0.17460 3	0.16473 3	0.15671 3
28	0.26385 3	0.23247 3	0.21009 3	0.19341 3	0.18052 3	0.17029 3	0.16199 3
29	0.27270 3	0.24021 3	0.21705 3	0.19978 3	0.18645 3	0.17586 3	0.16726 3
30	0.28154 3	0.24795 3	0.22401 3	0.20615 3	0.19237 3	0.18142 3	0.17253 3
31	0.29038 3	0.25570 3	0.23097 3	0.21253 3	0.19829 3	0.18698 3	0.17780 3
32	0.29923 3	0.26344 3	0.23792 3	0.21890 3	0.20421 3	0.19254 3	0.18307 3
33	0.30807 3	0.27118 3	0.24488 3	0.22527 3	0.21012 3	0.19810 3	0.18834 3
34	0.31691 3	0.27892 3	0.25183 3	0.23164 3	0.21604 3	0.20366 3	0.19360 3
35	0.32575 3	0.28666 3	0.25879 3	0.23800 3	0.22196 3	0.20921 3	0.19886 3
36	0.33458 3	0.29439 3	0.26574 3	0.24437 3	0.22787 3	0.21477 3	0.20413 3
37	0.34342 3	0.30213 3	0.27269 3	0.25074 3	0.23378 3	0.22032 3	0.20939 3
38	0.35226 3	0.30986 3	0.27964 3	0.25710 3	0.23970 3	0.22587 3	0.21465 3
39	0.36110 3	0.31760 3	0.28659 3	0.26347 3	0.24561 3	0.23143 3	0.21991 3
40	0.36993 3	0.32533 3	0.29354 3	0.26983 3	0.25152 3	0.23698 3	0.22516 3
41	0.37877 3	0.33307 3	0.30049 3	0.27619 3	0.25743 3	0.24253 3	0.23042 3
42	0.38760 3	0.34080 3	0.30744 3	0.28255 3	0.26334 3	0.24808 3	0.23568 3
43	0.39644 3	0.34853 3	0.31438 3	0.28892 3	0.26925 3	0.25362 3	0.24093 3
44	0.40527 3	0.35627 3	0.32133 3	0.29528 3	0.27515 3	0.25917 3	0.24619 3
45	0.41410 3	0.36400 3	0.32828 3	0.30164 3	0.28106 3	0.26472 3	0.25144 3
46	0.42294 3	0.37173 3	0.33522 3	0.30800 3	0.28697 3	0.27027 3	0.25670 3
47	0.43177 3	0.37946 3	0.34217 3	0.31436 3	0.29287 3	0.27581 3	0.26195 3
48	0.44060 3	0.38719 3	0.34911 3	0.32072 3	0.29878 3	0.28136 3	0.26720 3
49	0.44943 3	0.39492 3	0.35606 3	0.32707 3	0.30469 3	0.28690 3	0.27246 3
50	0.45827 3	0.40265 3	0.36300 3	0.33343 3	0.31059 3	0.29245 3	0.27771 3

(cont'd on next page)

TABLE 11-4 (cont'd)

1% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

m \ n	16	17	18	19	20	21	22
2	0.20255 2	0.19705 2	0.19235 2	0.18828 2	0.18472 2	0.18159 2	0.17881 2
3	0.26230 2	0.25477 2	0.24834 2	0.24279 2	0.23795 2	0.23369 2	0.22992 2
4	0.31897 2	0.30945 2	0.30133 2	0.29433 2	0.28823 2	0.28286 2	0.27811 2
5	0.37393 2	0.36244 2	0.35264 2	0.34418 2	0.33682 2	0.33036 2	0.32464 2
6	0.42781 2	0.41433 2	0.40285 2	0.39296 2	0.38434 2	0.37678 2	0.37008 2
7	0.48094 2	0.46548 2	0.45232 2	0.44097 2	0.43110 2	0.42244 2	0.41477 2
8	0.53353 2	0.51609 2	0.50124 2	0.48844 2	0.47731 2	0.46754 2	0.45889 2
9	0.58571 2	0.56629 2	0.54974 2	0.53549 2	0.52310 2	0.51221 2	0.50259 2
10	0.63758 2	0.61616 2	0.59793 2	0.58222 2	0.56855 2	0.55656 2	0.54595 2
11	0.68920 2	0.66579 2	0.64585 2	0.62868 2	0.61374 2	0.60063 2	0.58904 2
12	0.74062 2	0.71520 2	0.69356 2	0.67492 2	0.65871 2	0.64448 2	0.63190 2
13	0.79187 2	0.76445 2	0.74110 2	0.72100 2	0.70351 2	0.68816 2	0.67458 2
14	0.84298 2	0.81355 2	0.78850 2	0.76692 2	0.74815 2	0.73167 2	0.71711 2
15	0.89397 2	0.86254 2	0.83577 2	0.81272 2	0.79266 2	0.77506 2	0.75950 2
16	0.94487 2	0.91142 2	0.88294 2	0.85841 2	0.83707 2	0.81834 2	0.80177 2
17	0.99568 2	0.96021 2	0.93001 2	0.90400 2	0.88137 2	0.86151 2	0.84395 2
18	0.10464 2	0.10089 3	0.97701 2	0.94952 2	0.92560 2	0.90461 2	0.88604 2
19	0.10971 3	0.10576 3	0.10239 2	0.99496 2	0.96975 2	0.94762 2	0.92805 2
20	0.11477 3	0.11062 3	0.10708 3	0.10403 3	0.10138 3	0.99058 2	0.96999 2
21	0.11983 3	0.11547 3	0.11176 3	0.10857 3	0.10579 3	0.10335 3	0.10119 3
22	0.12488 3	0.12032 3	0.11644 3	0.11309 3	0.11019 3	0.10763 3	0.10537 3
23	0.12993 3	0.12517 3	0.12111 3	0.11762 3	0.11458 3	0.11191 3	0.10955 3
24	0.13497 3	0.13001 3	0.12578 3	0.12214 3	0.11897 3	0.11619 3	0.11372 3
25	0.14001 3	0.13484 3	0.13044 3	0.12665 3	0.12335 3	0.12046 3	0.11789 3
26	0.14505 3	0.13968 3	0.13511 3	0.13117 3	0.12774 3	0.12472 3	0.12206 3
27	0.15008 3	0.14451 3	0.13977 3	0.13568 3	0.13212 3	0.12899 3	0.12622 3
28	0.15512 3	0.14934 3	0.14442 3	0.14018 3	0.13649 3	0.13325 3	0.13038 3
29	0.16015 3	0.15417 3	0.14908 3	0.14469 3	0.14087 3	0.13751 3	0.13454 3
30	0.16518 3	0.15900 3	0.15373 3	0.14919 3	0.14524 3	0.14177 3	0.13870 3
31	0.17020 3	0.16382 3	0.15838 3	0.15369 3	0.14961 3	0.14602 3	0.14285 3
32	0.17523 3	0.16864 3	0.16303 3	0.15819 3	0.15397 3	0.15027 3	0.14700 3
33	0.18025 3	0.17346 3	0.16767 3	0.16268 3	0.15834 3	0.15452 3	0.15115 3
34	0.18528 3	0.17828 3	0.17232 3	0.16718 3	0.16270 3	0.15877 3	0.15529 3
35	0.19030 3	0.18310 3	0.17696 3	0.17167 3	0.16707 3	0.16302 3	0.15944 3
36	0.19532 3	0.18791 3	0.18160 3	0.17616 3	0.17143 3	0.16727 3	0.16358 3
37	0.20034 3	0.19273 3	0.18624 3	0.18065 3	0.17579 3	0.17151 3	0.16772 3
38	0.20535 3	0.19754 3	0.19088 3	0.18514 3	0.18014 3	0.17575 3	0.17186 3
39	0.21037 3	0.20235 3	0.19552 3	0.18963 2	0.18450 3	0.17999 3	0.17600 3
40	0.21539 3	0.20717 3	0.20016 3	0.19412 3	0.18886 3	0.18424 3	0.18014 3
41	0.22040 3	0.21198 3	0.20480 3	0.19860 3	0.19321 3	0.18847 3	0.18428 3
42	0.22542 3	0.21679 3	0.20943 3	0.20309 3	0.19757 3	0.19271 3	0.18842 3
43	0.23043 3	0.22160 3	0.21407 3	0.20757 3	0.20192 3	0.19695 3	0.19255 3
44	0.23544 3	0.22640 3	0.21870 3	0.21206 3	0.20627 3	0.20119 3	0.19668 3
45	0.24045 3	0.23121 3	0.22333 3	0.21654 3	0.21062 3	0.20542 3	0.20082 3
46	0.24547 3	0.23602 3	0.22797 3	0.22102 3	0.21497 3	0.20966 3	0.20495 3
47	0.25048 3	0.24083 3	0.23260 3	0.22550 3	0.21932 3	0.21389 3	0.20908 3
48	0.25549 3	0.24563 3	0.23723 3	0.22998 3	0.22367 3	0.21812 3	0.21321 3
49	0.26050 3	0.25044 3	0.24186 3	0.23446 3	0.22802 3	0.22236 3	0.21734 3
50	0.26551 3	0.25524 3	0.24649 3	0.23894 3	0.23237 3	0.22659 3	0.22147 3

(cont'd on next page)

TABLE 11-4 (cont'd)

1% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

m \ n	23	24	25	26	27	28	29
2	0.17633 2	0.17410 2	0.17209 2	0.17026 2	0.16859 2	0.16707 2	0.16567 2
3	0.22655 2	0.22353 2	0.22080 2	0.21833 2	0.21608 2	0.21402 2	0.21212 2
4	0.27388 2	0.27008 2	0.26665 2	0.26355 2	0.26072 2	0.25813 2	0.25576 2
5	0.31954 2	0.31497 2	0.31085 2	0.30711 2	0.30371 2	0.30060 2	0.29775 2
6	0.36412 2	0.35877 2	0.35396 2	0.34959 2	0.34562 2	0.34199 2	0.33866 2
7	0.40794 2	0.40182 2	0.39630 2	0.39131 2	0.38676 2	0.38261 2	0.37880 2
8	0.45120 2	0.44430 2	0.43808 2	0.43245 2	0.42733 2	0.42265 2	0.41836 2
9	0.49402 2	0.48634 2	0.47942 2	0.47315 2	0.46745 2	0.46225 2	0.45747 2
10	0.53650 2	0.52804 2	0.52042 2	0.51351 2	0.50723 2	0.50149 2	0.49623 2
11	0.57871 2	0.56946 2	0.56113 2	0.55358 2	0.54672 2	0.54044 2	0.53469 2
12	0.62070 2	0.61066 2	0.60161 2	0.59342 2	0.58597 2	0.57916 2	0.57292 2
13	0.66249 2	0.65166 2	0.64190 2	0.63307 2	0.62503 2	0.61768 2	0.61094 2
14	0.70413 2	0.69251 2	0.68203 2	0.67254 2	0.66391 2	0.65603 2	0.64880 2
15	0.74563 2	0.73321 2	0.72202 2	0.71188 2	0.70266 2	0.69423 2	0.68650 2
16	0.78702 2	0.77380 2	0.76188 2	0.75110 2	0.74128 2	0.73231 2	0.72408 2
17	0.82830 2	0.81428 2	0.80165 2	0.79020 2	0.77979 2	0.77028 2	0.76155 2
18	0.86950 2	0.85467 2	0.84132 2	0.82922 2	0.81821 2	0.80815 2	0.79892 2
19	0.91061 2	0.89499 2	0.88090 2	0.86815 2	0.85654 2	0.84593 2	0.83620 2
20	0.95166 2	0.93523 2	0.92042 2	0.90700 2	0.89480 2	0.88364 2	0.87340 2
21	0.99265 2	0.97541 2	0.95987 2	0.94580 2	0.93299 2	0.92128 2	0.91054 2
22	0.10336 3	0.10155 3	0.99926 2	0.98453 2	0.97112 2	0.95886 2	0.94761 2
23	0.10745 3	0.10556 3	0.10386 3	0.10232 3	0.10092 3	0.99638 2	0.98463 2
24	0.11153 3	0.10956 3	0.10779 3	0.10618 3	0.10472 3	0.10339 3	0.10216 3
25	0.11561 3	0.11356 3	0.11172 3	0.11004 3	0.10852 3	0.10713 3	0.10585 3
26	0.11969 3	0.11756 3	0.11564 3	0.11390 3	0.11231 3	0.11087 3	0.10954 3
27	0.12376 3	0.12155 3	0.11955 3	0.11775 3	0.11610 3	0.11460 3	0.11322 3
28	0.12783 3	0.12554 3	0.12347 3	0.12160 3	0.11989 3	0.11833 3	0.11690 3
29	0.13189 3	0.12952 3	0.12738 3	0.12544 3	0.12368 3	0.12206 3	0.12058 3
30	0.13596 3	0.13350 3	0.13129 3	0.12928 3	0.12746 3	0.12579 3	0.12425 3
31	0.14002 3	0.13748 3	0.13520 3	0.13312 3	0.13124 3	0.12951 3	0.12793 3
32	0.14408 3	0.14146 3	0.13910 3	0.13696 3	0.13501 3	0.13323 3	0.13159 3
33	0.14814 3	0.14544 3	0.14300 3	0.14079 3	0.13878 3	0.13695 3	0.13526 3
34	0.15219 3	0.14941 3	0.14690 3	0.14463 3	0.14256 3	0.14066 3	0.13892 3
35	0.15625 3	0.15338 3	0.15080 3	0.14846 3	0.14633 3	0.14438 3	0.14259 3
36	0.16030 3	0.15735 3	0.15470 3	0.15229 3	0.15009 3	0.14809 3	0.14625 3
37	0.16435 3	0.16132 3	0.15859 3	0.15611 3	0.15386 3	0.15180 3	0.14990 3
38	0.16840 3	0.16529 3	0.16248 3	0.15994 3	0.15762 3	0.15551 3	0.15356 3
39	0.17245 3	0.16925 3	0.16638 3	0.16376 3	0.16139 3	0.15921 3	0.15721 3
40	0.17649 3	0.17322 3	0.17027 3	0.16759 3	0.16515 3	0.16292 3	0.16087 2
41	0.18054 3	0.17718 3	0.17415 3	0.17141 3	0.16891 3	0.16662 3	0.16452 3
42	0.18458 3	0.18114 3	0.17804 3	0.17523 3	0.17267 3	0.17032 3	0.16817 3
43	0.18863 3	0.18511 3	0.18193 3	0.17905 3	0.17642 3	0.17402 3	0.17182 3
44	0.19267 2	0.18907 3	0.18582 3	0.18287 3	0.18018 3	0.17772 3	0.17547 3
45	0.19671 3	0.19303 3	0.18970 3	0.18668 3	0.18394 3	0.18142 3	0.17912 3
46	0.20075 3	0.19698 3	0.19358 3	0.19050 3	0.18769 3	0.18512 3	0.18276 3
47	0.20479 3	0.20094 3	0.19747 3	0.19432 3	0.19144 3	0.18882 3	0.18641 3
48	0.20883 3	0.20490 3	0.20135 3	0.19813 3	0.19520 3	0.19251 3	0.19005 3
49	0.21287 3	0.20885 3	0.20523 3	0.20194 3	0.19895 3	0.19621 3	0.19369 3
50	0.21691 3	0.21281 3	0.20911 3	0.20576 3	0.20270 3	0.19990 3	0.19734 3

(cont'd on next page)

TABLE 11-4 (cont'd)

1% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

m \ n	30	31	32	33	34	35	36
2	0.16438 2	0.16318 2	0.16207 2	0.16104 2	0.16008 2	0.15918 2	0.15834 2
3	0.21038 2	0.20877 2	0.20728 2	0.20589 2	0.20460 2	0.20339 2	0.20226 2
4	0.25358 2	0.25156 2	0.24969 2	0.24795 2	0.24633 2	0.24482 2	0.24340 2
5	0.29513 2	0.29271 2	0.29046 2	0.28837 2	0.28643 2	0.28462 2	0.28292 2
6	0.33560 2	0.33277 2	0.33015 2	0.32771 2	0.32544 2	0.32333 2	0.32135 2
7	0.37529 2	0.37205 2	0.36906 2	0.36627 2	0.36368 2	0.36126 2	0.35899 2
8	0.41441 2	0.41076 2	0.40738 2	0.40425 2	0.40133 2	0.39860 2	0.39605 2
9	0.45307 2	0.44902 2	0.44526 2	0.44177 2	0.43852 2	0.43548 2	0.43265 2
10	0.49138 2	0.48691 2	0.48277 2	0.47892 2	0.47534 2	0.47200 2	0.46887 2
11	0.52940 2	0.52451 2	0.51999 2	0.51578 2	0.51187 2	0.50822 2	0.50480 2
12	0.56717 2	0.56187 2	0.55696 2	0.55239 2	0.54815 2	0.54418 2	0.54047 2
13	0.60474 2	0.59902 2	0.59372 2	0.58880 2	0.58421 2	0.57993 2	0.57593 2
14	0.64214 2	0.63600 2	0.63031 2	0.62502 2	0.62010 2	0.61550 2	0.61121 2
15	0.67939 2	0.67282 2	0.66674 2	0.66109 2	0.65583 2	0.65092 2	0.64633 2
16	0.71651 2	0.70952 2	0.70304 2	0.69703 2	0.69143 2	0.68620 2	0.68131 2
17	0.75352 2	0.74610 2	0.73923 2	0.73285 2	0.72691 2	0.72136 2	0.71617 2
18	0.79042 2	0.78258 2	0.77531 2	0.76856 2	0.76228 2	0.75641 2	0.75092 2
19	0.82724 2	0.81897 2	0.81131 2	0.80419 2	0.79756 2	0.79137 2	0.78558 2
20	0.86398 2	0.85528 2	0.84722 2	0.83973 2	0.83276 2	0.82624 2	0.82015 2
21	0.90065 2	0.89152 2	0.88306 2	0.87520 2	0.86788 2	0.86104 2	0.85465 2
22	0.93726 2	0.92769 2	0.91883 2	0.91060 2	0.90293 2	0.89577 2	0.88907 2
23	0.97381 2	0.96381 2	0.95455 2	0.94594 2	0.93793 2	0.93044 2	0.92344 2
24	0.10103 3	0.99987 2	0.99021 2	0.98123 2	0.97287 2	0.96506 2	0.95775 2
25	0.10468 3	0.10359 3	0.10258 3	0.10165 3	0.10078 3	0.99962 2	0.99200 2
26	0.10832 3	0.10719 3	0.10614 3	0.10517 3	0.10426 3	0.10341 3	0.10262 3
27	0.11195 3	0.11078 3	0.10969 3	0.10868 3	0.10774 3	0.10686 3	0.10604 3
28	0.11559 3	0.11437 3	0.11324 3	0.11219 3	0.11122 3	0.11030 3	0.10945 3
29	0.11922 3	0.11795 3	0.11679 3	0.11570 3	0.11469 3	0.11374 3	0.11286 3
30	0.12284 3	0.12154 3	0.12033 3	0.11920 3	0.11816 3	0.11718 3	0.11626 3
31	0.12647 3	0.12512 3	0.12387 3	0.12271 3	0.12162 3	0.12061 3	0.11967 3
32	0.13009 3	0.12869 3	0.12740 3	0.12620 3	0.12509 3	0.12404 3	0.12307 3
33	0.13371 3	0.13227 3	0.13094 3	0.12970 3	0.12855 3	0.12747 3	0.12646 3
34	0.13732 3	0.13584 3	0.13447 3	0.13319 3	0.13201 3	0.13089 3	0.12986 3
35	0.14094 3	0.13941 3	0.13800 3	0.13669 3	0.13546 3	0.13432 3	0.13325 3
36	0.14455 3	0.14298 3	0.14153 3	0.14017 3	0.13891 3	0.13774 3	0.13664 3
37	0.14816 3	0.14655 3	0.14505 3	0.14366 3	0.14237 3	0.14116 3	0.14002 3
38	0.15177 3	0.15011 3	0.14858 3	0.14715 3	0.14582 3	0.14457 3	0.14341 3
39	0.15537 3	0.15367 3	0.15210 3	0.15063 3	0.14926 3	0.14799 3	0.14679 3
40	0.15898 3	0.15724 3	0.15562 3	0.15411 3	0.15271 3	0.15140 3	0.15017 2
41	0.16258 3	0.16079 3	0.15914 3	0.15759 3	0.15615 3	0.15481 3	0.15355 3
42	0.16619 3	0.16435 3	0.16265 3	0.16107 3	0.15960 3	0.15822 3	0.15693 3
43	0.16979 3	0.16791 3	0.16617 3	0.16455 3	0.16304 3	0.16163 3	0.16031 3
44	0.17339 3	0.17147 3	0.16968 3	0.16803 3	0.16648 3	0.16503 3	0.16368 3
45	0.17699 3	0.17502 3	0.17320 3	0.17150 3	0.16992 3	0.16844 3	0.16706 3
46	0.18059 3	0.17857 3	0.17671 3	0.17497 3	0.17336 3	0.17184 3	0.17043 3
47	0.18418 3	0.18213 3	0.18022 3	0.17845 3	0.17679 3	0.17525 3	0.17380 3
48	0.18778 3	0.18568 3	0.18373 3	0.18192 3	0.18023 3	0.17865 3	0.17717 3
49	0.19137 3	0.18923 3	0.18724 3	0.18539 3	0.18366 3	0.18205 3	0.18054 3
50	0.19497 3	0.19278 3	0.19075 3	0.18886 3	0.18710 3	0.18545 3	0.18391 3

(cont'd on next page)

TABLE 11-4 (cont'd)

1% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

n / m	37	38	39	40	41	42	43
2	0.15756 2	0.15681 2	0.15611 2	0.15545 2	0.15483 2	0.15424 2	0.15368 2
3	0.20120 2	0.20020 2	0.19926 2	0.19838 2	0.19754 2	0.19675 2	0.19600 2
4	0.24208 2	0.24083 2	0.23966 2	0.23855 2	0.23750 2	0.23652 2	0.23558 2
5	0.28133 2	0.27983 2	0.27843 2	0.27710 2	0.27584 2	0.27466 2	0.27353 2
6	0.31949 2	0.31775 2	0.31610 2	0.31456 2	0.31309 2	0.31171 2	0.31040 2
7	0.35687 2	0.35488 2	0.35300 2	0.35123 2	0.34956 2	0.34798 2	0.34648 2
8	0.39366 2	0.39141 2	0.38930 2	0.38731 2	0.38543 2	0.38365 2	0.38196 2
9	0.42999 2	0.42749 2	0.42514 2	0.42292 2	0.42083 2	0.41885 2	0.41697 2
10	0.46594 2	0.46319 2	0.46060 2	0.45816 2	0.45585 2	0.45367 2	0.45160 2
11	0.50160 2	0.49859 2	0.49576 2	0.49309 2	0.49057 2	0.48818 2	0.48592 2
12	0.53699 2	0.53373 2	0.53065 2	0.52776 2	0.52502 2	0.52243 2	0.51998 2
13	0.57218 2	0.56865 2	0.56534 2	0.56221 2	0.55926 2	0.55646 2	0.55382 2
14	0.60718 2	0.60339 2	0.59983 2	0.59647 2	0.59330 2	0.59030 2	0.58746 2
15	0.64202 2	0.63797 2	0.63416 2	0.63057 2	0.62718 2	0.62398 2	0.62094 2
16	0.67672 2	0.67241 2	0.66836 2	0.66453 2	0.66092 2	0.65751 2	0.65427 2
17	0.71130 2	0.70673 2	0.70242 2	0.69837 2	0.69453 2	0.69091 2	0.68747 2
18	0.74577 2	0.74094 3	0.73638 2	0.73209 2	0.72804 2	0.72420 2	0.72057 2
19	0.78015 2	0.77505 2	0.77024 2	0.76571 2	0.76144 2	0.75739 2	0.75355 2
20	0.81444 2	0.80907 2	0.80401 2	0.79925 2	0.79475 2	0.79049 2	0.78645 2
21	0.84865 2	0.84301 2	0.83771 2	0.83270 2	0.82798 2	0.82351 2	0.81927 2
22	0.88279 2	0.87689 2	0.87133 2	0.86609 2	0.86113 2	0.85645 2	0.85201 2
23	0.91687 2	0.91070 2	0.90489 2	0.89940 2	0.89423 2	0.88933 2	0.88468 2
24	0.95089 2	0.94445 2	0.93838 2	0.93266 2	0.92726 2	0.92214 2	0.91729 2
25	0.98486 2	0.97815 2	0.97183 2	0.96586 2	0.96023 2	0.95490 2	0.94985 2
26	0.10188 3	0.10118 3	0.10052 3	0.99902 2	0.99316 3	0.98761 3	0.98235 3
27	0.10527 3	0.10454 3	0.10386 3	0.10321 3	0.10260 3	0.10203 3	0.10148 3
28	0.10865 3	0.10790 3	0.10719 3	0.10652 3	0.10589 3	0.10529 3	0.10472 3
29	0.11203 3	0.11125 3	0.11051 3	0.10982 3	0.10917 3	0.10855 3	0.10796 3
30	0.11541 3	0.11460 3	0.11384 3	0.11312 3	0.11244 3	0.11180 3	0.11119 3
31	0.11878 3	0.11794 3	0.11716 3	0.11642 3	0.11571 3	0.11505 3	0.11442 3
32	0.12215 3	0.12129 3	0.12047 3	0.11971 3	0.11898 3	0.11830 3	0.11765 3
33	0.12552 3	0.12463 3	0.12379 3	0.12300 3	0.12225 3	0.12154 3	0.12087 3
34	0.12888 3	0.12796 3	0.12710 3	0.12628 3	0.12551 3	0.12478 3	0.12409 3
35	0.13224 3	0.13130 3	0.13041 3	0.12957 3	0.12877 3	0.12802 3	0.12731 3
36	0.13560 3	0.13463 3	0.13371 3	0.13285 3	0.13203 3	0.13126 3	0.13052 3
37	0.13896 3	0.13796 3	0.13702 3	0.13613 3	0.13529 3	0.13449 3	0.13374 3
38	0.14231 3	0.14129 3	0.14032 3	0.13940 3	0.13854 3	0.13772 3	0.13695 3
39	0.14567 3	0.14461 3	0.14362 3	0.14268 2	0.14179 3	0.14095 3	0.14016 3
40	0.14902 3	0.14794 3	0.14692 3	0.14595 3	0.14504 3	0.14418 3	0.14336 3
41	0.15237 3	0.15126 3	0.15021 3	0.14922 3	0.14829 3	0.14741 3	0.14657 3
42	0.15572 3	0.15458 3	0.15351 3	0.15249 3	0.15154 3	0.15063 3	0.14977 3
43	0.15907 3	0.15790 3	0.15680 3	0.15576 3	0.15478 3	0.15385 3	0.15297 3
44	0.16241 3	0.16122 3	0.16009 3	0.15903 3	0.15803 3	0.15707 3	0.15617 3
45	0.16576 3	0.16453 3	0.16338 3	0.16229 3	0.16127 3	0.16029 3	0.15937 3
46	0.16910 3	0.16785 3	0.16667 3	0.16556 3	0.16451 3	0.16351 3	0.16257 3
47	0.17244 3	0.17116 3	0.16996 3	0.16882 3	0.16775 3	0.16673 3	0.16576 3
48	0.17578 3	0.17448 3	0.17324 3	0.17208 3	0.17098 3	0.16994 3	0.16896 3
49	0.17912 3	0.17779 3	0.17653 3	0.17534 3	0.17422 3	0.17316 3	0.17215 3
50	0.18246 3	0.18110 3	0.17981 3	0.17860 3	0.17746 3	0.17637 3	0.17534 3

(cont'd on next page)

TABLE 11-4 (cont'd)

1% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

m \ n	44	45	46	47	48	49	50
2	0.15315 2	0.15265 2	0.15217 2	0.15171 2	0.15127 2	0.15085 2	0.15046 2
3	0.19528 2	0.19461 2	0.19396 2	0.19335 2	0.19276 2	0.19221 2	0.19167 2
4	0.23469 2	0.23384 2	0.23304 2	0.23227 2	0.23154 2	0.23085 2	0.23018 2
5	0.27247 2	0.27146 2	0.27049 2	0.26958 2	0.26870 2	0.26787 2	0.26707 2
6	0.30916 2	0.30798 2	0.30686 2	0.30579 2	0.30477 2	0.30379 2	0.30286 2
7	0.34506 2	0.34371 2	0.34243 2	0.34121 2	0.34004 2	0.33893 2	0.33787 2
8	0.38036 2	0.37884 2	0.37740 2	0.37602 2	0.37471 2	0.37346 2	0.37226 2
9	0.41519 2	0.41350 2	0.41190 2	0.41037 2	0.40891 2	0.40751 2	0.40618 2
10	0.44964 2	0.44778 2	0.44601 2	0.44432 2	0.44272 2	0.44118 2	0.43972 2
11	0.48378 2	0.48175 2	0.47981 2	0.47797 2	0.47621 2	0.47453 2	0.47293 2
12	0.51765 2	0.51544 2	0.51334 2	0.51134 2	0.50944 2	0.50761 2	0.50588 2
13	0.55130 2	0.54892 2	0.54665 2	0.54449 2	0.54243 2	0.54047 2	0.53859 2
14	0.58476 2	0.58220 2	0.57977 2	0.57745 2	0.57524 2	0.57312 2	0.57111 2
15	0.61805 2	0.61532 2	0.61271 2	0.61023 2	0.60787 2	0.60561 2	0.60345 2
16	0.65120 2	0.64828 2	0.64551 2	0.64287 2	0.64035 2	0.63794 2	0.63565 2
17	0.68422 2	0.68112 2	0.67818 2	0.67537 2	0.67270 2	0.67015 2	0.66771 2
18	0.71712 2	0.71384 2	0.71073 2	0.70776 2	0.70493 2	0.70223 2	0.69965 2
19	0.74992 2	0.74646 2	0.74317 2	0.74004 2	0.73705 2	0.73420 2	0.73148 2
20	0.78262 2	0.77899 2	0.77552 2	0.77223 2	0.76909 2	0.76608 2	0.76322 2
21	0.81525 2	0.81143 2	0.80779 2	0.80433 2	0.80103 2	0.79788 2	0.79487 2
22	0.84780 2	0.84379 2	0.83998 2	0.83636 2	0.83290 2	0.82960 2	0.82644 2
23	0.88028 2	0.87609 2	0.87211 2	0.86831 2	0.86470 2	0.86124 2	0.85794 2
24	0.91270 2	0.90832 2	0.90417 2	0.90021 2	0.89643 2	0.89282 2	0.88937 2
25	0.94506 2	0.94050 2	0.93617 2	0.93204 2	0.92810 2	0.92434 2	0.92075 2
26	0.97736 2	0.97262 2	0.96811 2	0.96382 2	0.95972 2	0.95580 2	0.95206 2
27	0.10096 3	0.10047 3	0.10000 3	0.99555 2	0.99129 2	0.98722 2	0.98333 2
28	0.10418 3	0.10367 3	0.10319 3	0.10272 3	0.10228 3	0.10186 3	0.10145 3
29	0.10740 3	0.10687 3	0.10637 3	0.10589 3	0.10543 3	0.10499 3	0.10457 3
30	0.11061 3	0.11007 3	0.10954 3	0.10905 3	0.10857 3	0.10812 3	0.10769 3
31	0.11382 3	0.11326 3	0.11272 3	0.11220 3	0.11171 3	0.11124 3	0.11080 3
32	0.11703 3	0.11645 3	0.11589 3	0.11536 3	0.11485 3	0.11436 3	0.11390 3
33	0.12024 3	0.11963 3	0.11905 3	0.11851 3	0.11798 3	0.11748 3	0.11700 3
34	0.12344 3	0.12281 3	0.12222 3	0.12165 3	0.12111 3	0.12060 3	0.12010 3
35	0.12663 3	0.12599 3	0.12538 3	0.12480 3	0.12424 3	0.12371 3	0.12320 3
36	0.12983 3	0.12917 3	0.12854 3	0.12794 3	0.12736 3	0.12682 3	0.12629 3
37	0.13302 3	0.13234 3	0.13169 3	0.13108 3	0.13049 3	0.12992 3	0.12939 3
38	0.13621 3	0.13551 3	0.13485 3	0.13421 3	0.13361 3	0.13303 3	0.13247 3
39	0.13940 3	0.13868 3	0.13800 3	0.13735 3	0.13672 3	0.13613 3	0.13556 3
40	0.14259 3	0.14185 3	0.14115 3	0.14048 3	0.13984 3	0.13923 3	0.13864 3
41	0.14577 3	0.14502 3	0.14429 3	0.14361 3	0.14295 3	0.14233 3	0.14173 3
42	0.14895 3	0.14818 3	0.14744 3	0.14674 3	0.14606 3	0.14542 3	0.14481 3
43	0.15214 3	0.15134 3	0.15058 3	0.14986 3	0.14917 3	0.14852 3	0.14789 3
44	0.15532 2	0.15450 3	0.15373 3	0.15299 3	0.15228 3	0.15161 3	0.15096 3
45	0.15849 3	0.15766 3	0.15687 3	0.15611 3	0.15539 3	0.15470 3	0.15404 3
46	0.16167 3	0.16082 3	0.16001 3	0.15923 3	0.15849 3	0.15779 3	0.15711 3
47	0.16485 3	0.16397 3	0.16314 3	0.16235 3	0.16160 3	0.16087 3	0.16018 3
48	0.16802 3	0.16713 3	0.16628 3	0.16547 3	0.16470 3	0.16396 3	0.16325 3
49	0.17119 3	0.17028 3	0.16941 3	0.16859 3	0.16780 3	0.16704 3	0.16632 3
50	0.17436 3	0.17343 3	0.17255 3	0.17170 3	0.17090 3	0.17013 3	0.16939 3

(cont'd on next page)

TABLE 11-4 (cont'd)

5% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

n / m	2	3	4	5	6	7	8
2	0.19718 4	0.11686 3	0.47998 2	0.31278 2	0.24350 2	0.20668 2	0.18414 2
3	0.31978 4	0.17675 3	0.69937 2	0.44593 2	0.34234 2	0.28781 2	0.25464 2
4	0.44390 2	0.23664 3	0.91673 2	0.57687 2	0.43899 2	0.36677 2	0.32301 2
5	0.56865 4	0.29654 3	0.11331 3	0.70674 2	0.53454 2	0.44463 2	0.39027 2
6	0.69371 4	0.35664 3	0.13490 3	0.83602 2	0.62947 2	0.52185 2	0.45687 2
7	0.81895 4	0.41635 3	0.15646 3	0.96492 2	0.72400 2	0.59865 2	0.52304 2
8	0.94429 4	0.47625 3	0.17800 3	0.10936 3	0.81826 2	0.67517 2	0.58891 2
9	0.10697 5	0.53615 3	0.19952 3	0.12221 3	0.91234 2	0.75148 2	0.65458 2
10	0.11952 5	0.59606 3	0.22104 3	0.13504 3	0.10063 3	0.82765 2	0.72008 2
11	0.13207 5	0.65596 3	0.24254 3	0.14787 3	0.11001 3	0.90371 2	0.78547 2
12	0.14462 5	0.71586 3	0.26404 3	0.16069 3	0.11939 3	0.97967 2	0.85075 2
13	0.15717 5	0.77577 3	0.28554 3	0.17351 3	0.12876 3	0.10556 3	0.91597 2
14	0.16973 5	0.83567 3	0.30703 3	0.18632 3	0.13812 3	0.11314 3	0.98112 2
15	0.18229 5	0.89557 3	0.32852 3	0.19912 3	0.14748 3	0.12072 3	0.10462 3
16	0.19485 5	0.95548 3	0.35001 3	0.21193 3	0.15684 3	0.12830 3	0.11113 3
17	0.20741 5	0.10154 4	0.37149 3	0.22473 3	0.16619 3	0.13587 3	0.11763 3
18	0.21997 5	0.10753 4	0.39298 3	0.23753 3	0.17554 3	0.14344 3	0.12413 3
19	0.23253 5	0.11352 4	0.41446 3	0.25033 3	0.18489 3	0.15100 3	0.13063 3
20	0.24509 5	0.11951 4	0.43594 3	0.26312 3	0.19424 3	0.15857 3	0.13712 3
21	0.25765 5	0.12550 4	0.45742 3	0.27592 3	0.20358 3	0.16613 3	0.14361 3
22	0.27021 5	0.13149 4	0.47890 3	0.28871 3	0.21293 3	0.17369 3	0.15010 3
23	0.28277 5	0.13748 4	0.50038 3	0.30150 3	0.22227 3	0.18125 3	0.15659 3
24	0.29534 5	0.14347 4	0.52186 3	0.31429 3	0.23161 3	0.18881 3	0.16308 3
25	0.30790 5	0.14946 4	0.54334 3	0.32708 3	0.24095 3	0.19637 3	0.16957 3
26	0.32046 5	0.15545 4	0.56481 3	0.33987 3	0.25029 3	0.20393 3	0.17605 3
27	0.33303 5	0.16144 4	0.58629 3	0.35266 3	0.25963 3	0.21149 3	0.18254 3
28	0.34559 5	0.16743 4	0.60777 3	0.36545 3	0.26897 3	0.21904 3	0.18902 3
29	0.35815 5	0.17342 4	0.62924 3	0.37824 3	0.27831 3	0.22660 3	0.19550 3
30	0.37072 5	0.17941 4	0.65072 3	0.39103 3	0.28765 3	0.23415 3	0.20198 3
31	0.38328 5	0.18540 4	0.67219 3	0.40382 3	0.29698 3	0.24170 3	0.20846 3
32	0.39584 5	0.19139 4	0.69367 3	0.41660 3	0.30632 3	0.24926 3	0.21495 3
33	0.40841 5	0.19739 4	0.71514 3	0.42939 3	0.31566 3	0.25681 3	0.22143 3
34	0.42097 5	0.20338 4	0.73662 3	0.44218 3	0.32499 3	0.26436 3	0.22791 3
35	0.43354 5	0.20937 4	0.75809 3	0.45496 3	0.33433 3	0.27191 3	0.23439 3
36	0.44610 5	0.21536 4	0.77956 3	0.46775 3	0.34366 3	0.27947 3	0.24086 3
37	0.45866 5	0.22135 4	0.80104 3	0.48053 3	0.35300 3	0.28702 3	0.24734 3
38	0.47123 5	0.22734 4	0.82251 3	0.49332 3	0.36233 3	0.29457 3	0.25382 3
39	0.48379 5	0.23333 4	0.84399 3	0.50611 3	0.37167 3	0.30212 3	0.26030 3
40	0.49636 5	0.23932 4	0.86546 3	0.51889 3	0.38100 3	0.30967 3	0.26678 3
41	0.50892 5	0.24531 4	0.88693 3	0.53168 3	0.39034 3	0.31722 3	0.27325 3
42	0.52149 5	0.25130 4	0.90841 3	0.54446 3	0.39967 3	0.32477 3	0.27973 3
43	0.53405 5	0.25729 4	0.92988 3	0.55724 3	0.40900 3	0.33232 3	0.28621 3
44	0.54662 5	0.26328 4	0.95135 3	0.57003 3	0.41834 3	0.33987 3	0.29268 3
45	0.55918 5	0.26927 4	0.97282 3	0.58281 3	0.42767 3	0.34742 3	0.29916 3
46	0.57175 5	0.27526 4	0.99430 3	0.59560 3	0.43700 3	0.35497 3	0.30564 3
47	0.58431 5	0.28125 4	0.10158 4	0.60838 3	0.44634 3	0.36251 3	0.31211 3
48	0.59688 5	0.28724 4	0.10372 4	0.62117 3	0.45567 3	0.37006 3	0.31859 3
49	0.60944 5	0.29323 4	0.10587 4	0.63395 3	0.46500 3	0.37761 3	0.32506 3
50	0.62201 5	0.29922 4	0.10802 4	0.64673 3	0.47434 3	0.38516 3	0.33154 3

(cont'd on next page)

TABLE 11-4 (cont'd)

5% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

m \ n	9	10	11	12	13	14	15
2	0.16901 2	0.15819 2	0.15009 2	0.14380 2	0.13879 2	0.13470 2	0.13130 2
3	0.23250 2	0.21673 2	0.20496 2	0.19585 2	0.18861 2	0.18270 2	0.17781 2
4	0.29388 2	0.27318 2	0.25776 2	0.24584 2	0.23637 2	0.22867 2	0.22229 2
5	0.35414 2	0.32851 2	0.30943 2	0.29471 2	0.28302 2	0.27351 2	0.26564 2
6	0.41373 2	0.38316 2	0.36042 2	0.34288 2	0.32896 2	0.31765 2	0.30828 2
7	0.47289 2	0.43736 2	0.41095 2	0.39058 2	0.37442 2	0.36129 2	0.35043 2
8	0.53173 2	0.49124 2	0.46115 2	0.43795 2	0.41954 2	0.40460 2	0.39222 2
9	0.59036 2	0.54489 2	0.51111 2	0.48507 2	0.46441 2	0.44764 2	0.43375 2
10	0.64882 2	0.59837 2	0.56089 2	0.53200 2	0.50909 2	0.49048 2	0.47508 2
11	0.70714 2	0.65171 2	0.61053 2	0.57879 2	0.55361 2	0.53316 2	0.51624 2
12	0.76537 2	0.70494 2	0.66005 2	0.62546 2	0.59801 2	0.57572 2	0.55727 2
13	0.82352 2	0.75809 2	0.70949 2	0.67203 2	0.64231 2	0.61818 2	0.59820 2
14	0.88160 2	0.81117 2	0.75885 2	0.71852 2	0.68653 2	0.66054 2	0.63904 2
15	0.93962 2	0.86418 2	0.80814 2	0.76495 2	0.73068 2	0.70284 2	0.67980 2
16	0.99760 2	0.91715 2	0.85739 2	0.81132 2	0.77477 2	0.74508 2	0.72050 2
17	0.10555 3	0.97008 2	0.90659 2	0.85764 2	0.81881 2	0.78726 2	0.76114 2
18	0.11134 3	0.10230 2	0.95575 2	0.90393 2	0.86280 2	0.82940 2	0.80173 2
19	0.11713 3	0.10758 3	0.10049 3	0.95018 2	0.90676 2	0.87149 2	0.84229 2
20	0.12292 3	0.11287 3	0.10540 3	0.99639 2	0.95069 2	0.91356 2	0.88280 2
21	0.12870 3	0.11815 3	0.11030 3	0.10426 3	0.99459 2	0.95559 2	0.92329 2
22	0.13448 3	0.12343 3	0.11521 3	0.10887 3	0.10385 3	0.99759 2	0.96375 2
23	0.14026 3	0.12870 3	0.12011 3	0.11349 3	0.10823 3	0.10396 3	0.10042 3
24	0.14604 3	0.13398 3	0.12501 3	0.11810 3	0.11261 3	0.10815 3	0.10446 3
25	0.15182 3	0.13925 3	0.12991 3	0.12271 3	0.11699 3	0.11235 3	0.10850 3
26	0.15759 3	0.14452 3	0.13481 3	0.12732 3	0.12137 3	0.11654 3	0.11254 3
27	0.16337 3	0.14980 3	0.13971 3	0.13193 3	0.12575 3	0.12073 3	0.11657 3
28	0.16914 3	0.15507 3	0.14461 3	0.13654 3	0.13013 3	0.12492 3	0.12060 3
29	0.17491 3	0.16033 3	0.14950 3	0.14114 3	0.13450 3	0.12911 3	0.12464 3
30	0.18068 3	0.16560 3	0.15439 3	0.14575 3	0.13888 3	0.13330 3	0.12867 3
31	0.18645 3	0.17087 3	0.15929 3	0.15035 3	0.14325 3	0.13748 3	0.13270 3
32	0.19222 3	0.17614 3	0.16418 3	0.15495 3	0.14763 3	0.14167 3	0.13673 3
33	0.19799 3	0.18140 3	0.16907 3	0.15956 3	0.15200 3	0.14585 3	0.14076 3
34	0.20376 3	0.18667 3	0.17396 3	0.16416 3	0.15637 3	0.15003 3	0.14478 3
35	0.20953 3	0.19194 3	0.17885 3	0.16876 3	0.16074 3	0.15422 3	0.14881 3
36	0.21530 3	0.19720 3	0.18374 3	0.17336 3	0.16511 3	0.15840 3	0.15284 3
37	0.22107 3	0.20246 3	0.18863 3	0.17796 3	0.16948 3	0.16258 3	0.15686 3
38	0.22684 3	0.20773 3	0.19352 3	0.18256 3	0.17384 3	0.16676 3	0.16088 3
39	0.23260 3	0.21299 3	0.19841 3	0.18715 3	0.17821 3	0.17094 3	0.16491 3
40	0.23837 3	0.21825 3	0.20330 3	0.19175 3	0.18258 3	0.17512 3	0.16893 2
41	0.24414 3	0.22352 3	0.20818 3	0.19635 3	0.18695 3	0.17930 3	0.17295 3
42	0.24990 3	0.22878 3	0.21307 3	0.20095 3	0.19131 3	0.18348 3	0.17698 3
43	0.25567 3	0.23404 3	0.21796 3	0.20554 3	0.19568 3	0.18765 3	0.18100 3
44	0.26143 2	0.23930 3	0.22284 3	0.21014 3	0.20004 3	0.19183 3	0.18502 3
45	0.26720 3	0.24456 3	0.22773 3	0.21474 3	0.20441 3	0.19601 3	0.18904 3
46	0.27296 3	0.24982 3	0.23262 3	0.21933 3	0.20877 3	0.20018 3	0.19306 3
47	0.27873 3	0.25509 3	0.23750 3	0.22393 3	0.21314 3	0.20436 3	0.19708 3
48	0.28449 3	0.26035 3	0.24239 3	0.22852 3	0.21750 3	0.20854 3	0.20110 3
49	0.29026 3	0.26561 3	0.24727 3	0.23312 3	0.22187 3	0.21271 3	0.20512 3
50	0.29602 3	0.27087 3	0.25216 3	0.23771 3	0.22623 3	0.21689 3	0.20914 3

(cont'd on next page)

TABLE 11-4 (cont'd)

5% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

n / m	16	17	18	19	20	21	22
2	0.12843 2	0.12598 2	0.12386 2	0.12201 2	0.12038 2	0.11894 2	0.11765 2
3	0.17369 2	0.17016 2	0.16712 2	0.16447 2	0.16214 2	0.16007 2	0.15822 2
4	0.21692 2	0.21233 2	0.20838 2	0.20493 2	0.20190 2	0.19921 2	0.19682 2
5	0.25902 2	0.25337 2	0.24850 2	0.24425 2	0.24052 2	0.23722 2	0.23427 2
6	0.30040 2	0.29369 2	0.28789 2	0.28285 2	0.27841 2	0.27449 2	0.27099 2
7	0.34129 2	0.33350 2	0.32678 2	0.32093 2	0.31579 2	0.31124 2	0.30718 2
8	0.38182 2	0.37295 2	0.36530 2	0.35864 2	0.35279 2	0.34761 2	0.34299 2
9	0.42208 2	0.41213 2	0.40355 2	0.39607 2	0.38951 2	0.38369 2	0.37851 2
10	0.46213 2	0.45109 2	0.44157 2	0.43328 2	0.42600 2	0.41955 2	0.41380 2
11	0.50201 2	0.48988 2	0.47942 2	0.47031 2	0.46231 2	0.45522 2	0.44890 2
12	0.54176 2	0.52854 2	0.51713 2	0.50720 2	0.49847 2	0.49074 2	0.48385 2
13	0.58140 2	0.56708 2	0.55473 2	0.54397 2	0.53451 2	0.52614 2	0.51867 2
14	0.62095 2	0.60552 2	0.59222 2	0.58063 2	0.57045 2	0.56143 2	0.55339 2
15	0.66042 2	0.64389 2	0.62964 2	0.61722 2	0.60630 2	0.59663 2	0.58801 2
16	0.69982 2	0.68219 2	0.66698 2	0.65373 2	0.64208 2	0.63176 2	0.62256 2
17	0.73916 2	0.72042 2	0.70426 2	0.69017 2	0.67779 2	0.66682 2	0.65704 2
18	0.77846 2	0.75861 3	0.74149 2	0.72657 2	0.71345 2	0.70182 2	0.69146 2
19	0.81771 2	0.79675 2	0.77867 2	0.76291 2	0.74905 2	0.73678 2	0.72582 2
20	0.85693 2	0.83486 2	0.81581 2	0.79921 2	0.78462 2	0.77168 2	0.76014 2
21	0.89611 2	0.87292 2	0.85292 2	0.83548 2	0.82014 2	0.80655 2	0.79442 2
22	0.93526 2	0.91096 2	0.88999 2	0.87171 2	0.85563 2	0.84138 2	0.82867 2
23	0.97439 2	0.94897 2	0.92703 2	0.90791 2	0.89109 2	0.87618 2	0.86288 2
24	0.10135 3	0.98696 2	0.96405 2	0.94408 2	0.92652 2	0.91095 2	0.89706 2
25	0.10526 3	0.10249 3	0.10010 3	0.98023 2	0.96192 2	0.94570 2	0.93121 2
26	0.10916 3	0.10629 3	0.10380 3	0.10164 3	0.99731 2	0.98042 2	0.96534 2
27	0.11307 3	0.11008 3	0.10750 3	0.10525 3	0.10327 3	0.10151 3	0.99945 2
28	0.11697 3	0.11387 3	0.11119 3	0.10886 3	0.10680 3	0.10498 3	0.10335 3
29	0.12087 3	0.11766 3	0.11488 3	0.11246 3	0.11033 3	0.10845 3	0.10676 3
30	0.12477 3	0.12145 3	0.11857 3	0.11607 3	0.11386 3	0.11191 3	0.11017 3
31	0.12867 3	0.12523 3	0.12226 3	0.11967 3	0.11739 3	0.11537 3	0.11357 3
32	0.13257 3	0.12902 3	0.12595 3	0.12328 3	0.12092 3	0.11884 3	0.11697 3
33	0.13647 3	0.13280 3	0.12964 3	0.12688 3	0.12445 3	0.12230 3	0.12037 3
34	0.14036 3	0.13659 3	0.13332 3	0.13048 3	0.12798 3	0.12575 3	0.12377 3
35	0.14426 3	0.14037 3	0.13701 3	0.13408 3	0.13150 3	0.12921 3	0.12717 3
36	0.14815 3	0.14415 3	0.14069 3	0.13768 3	0.13502 3	0.13267 3	0.13057 3
37	0.15204 3	0.14793 3	0.14438 3	0.14128 3	0.13855 3	0.13613 3	0.13396 3
38	0.15594 3	0.15171 3	0.14806 3	0.14487 3	0.14207 3	0.13958 3	0.13736 3
39	0.15983 3	0.15549 3	0.15174 3	0.14847 2	0.14559 3	0.14303 3	0.14075 3
40	0.16372 3	0.15927 3	0.15542 3	0.15207 3	0.14911 3	0.14649 3	0.14415 3
41	0.16761 3	0.16305 3	0.15910 3	0.15566 3	0.15263 3	0.14994 3	0.14754 3
42	0.17150 3	0.16682 3	0.16278 3	0.15925 3	0.15615 3	0.15339 3	0.15093 3
43	0.17539 3	0.17060 3	0.16646 3	0.16285 3	0.15967 3	0.15684 3	0.15432 3
44	0.17928 3	0.17438 3	0.17014 3	0.16644 3	0.16319 3	0.16030 3	0.15771 3
45	0.18317 3	0.17815 3	0.17382 3	0.17003 3	0.16670 3	0.16375 3	0.16110 3
46	0.18706 3	0.18193 3	0.17750 3	0.17363 3	0.17022 3	0.16720 3	0.16449 3
47	0.19094 3	0.18570 3	0.18117 3	0.17722 3	0.17374 3	0.17064 3	0.16788 3
48	0.19483 3	0.18948 3	0.18485 3	0.18081 3	0.17725 3	0.17409 3	0.17127 3
49	0.19872 3	0.19325 3	0.18853 3	0.18440 3	0.18077 3	0.17754 3	0.17466 3
50	0.20261 3	0.19703 3	0.19220 3	0.18799 3	0.18428 3	0.18099 3	0.17805 3

(cont'd on next page)

TABLE 11-4 (cont'd)

5% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

n / m	23	24	25	26	27	28	29
2	0.11649 2	0.11544 2	0.11449 2	0.11362 2	0.11283 2	0.11210 2	0.11142 2
3	0.15657 2	0.15507 2	0.15371 2	0.15248 2	0.15134 2	0.15030 2	0.14935 2
4	0.19467 2	0.19273 2	0.19097 2	0.18936 2	0.18790 2	0.18655 2	0.18531 2
5	0.23163 2	0.22924 2	0.22708 2	0.22511 2	0.22330 2	0.22165 2	0.22013 2
6	0.26784 2	0.26501 2	0.26244 2	0.26010 2	0.25796 2	0.25600 2	0.25419 2
7	0.30354 2	0.30025 2	0.29728 2	0.29457 2	0.29209 2	0.28981 2	0.28771 2
8	0.33885 2	0.33511 2	0.33172 2	0.32864 2	0.32581 2	0.32322 2	0.32084 2
9	0.37386 2	0.36967 2	0.36586 2	0.36240 2	0.35923 2	0.35633 2	0.35365 2
10	0.40864 2	0.40399 2	0.39977 2	0.39593 2	0.39241 2	0.38919 2	0.38622 2
11	0.44323 2	0.43811 2	0.43348 2	0.42925 2	0.42539 2	0.42185 2	0.41858 2
12	0.47766 2	0.47209 2	0.46703 2	0.46242 2	0.45821 2	0.45434 2	0.45078 2
13	0.51197 2	0.50593 2	0.50045 2	0.49546 2	0.49089 2	0.48670 2	0.48284 2
14	0.54617 2	0.53966 2	0.53375 2	0.52837 2	0.52346 2	0.51894 2	0.51478 2
15	0.58027 2	0.57329 2	0.56696 2	0.56120 2	0.55592 2	0.55108 2	0.54661 2
16	0.61430 2	0.60685 2	0.60009 2	0.59393 2	0.58830 2	0.58313 2	0.57836 2
17	0.64826 2	0.64033 2	0.63314 2	0.62660 2	0.62061 2	0.61510 2	0.61004 2
18	0.68215 2	0.67375 3	0.66613 2	0.65919 2	0.65284 2	0.64701 2	0.64164 2
19	0.71599 2	0.70712 2	0.69907 2	0.69174 2	0.68502 2	0.67886 2	0.67318 2
20	0.74979 2	0.74044 2	0.73195 2	0.72423 2	0.71715 2	0.71066 2	0.70467 2
21	0.78354 2	0.77371 2	0.76480 2	0.75667 2	0.74924 2	0.74241 2	0.73611 2
22	0.81725 2	0.80695 2	0.79760 2	0.78908 2	0.78128 2	0.77411 2	0.76751 2
23	0.85093 2	0.84015 2	0.83036 2	0.82144 2	0.81328 2	0.80578 2	0.79887 2
24	0.88458 2	0.87332 2	0.86310 2	0.85378 2	0.84525 2	0.83742 2	0.83019 2
25	0.91820 2	0.90646 2	0.89580 2	0.88608 2	0.87719 2	0.86902 2	0.86149 2
26	0.95180 2	0.93957 2	0.92848 2	0.91836 2	0.90910 2	0.90059 2	0.89275 2
27	0.98538 2	0.97267 2	0.96113 2	0.95062 2	0.94099 2	0.93214 2	0.92398 2
28	0.10189 3	0.10057 3	0.99376 2	0.98285 2	0.97285 2	0.96367 2	0.95520 2
29	0.10525 3	0.10388 3	0.10264 3	0.10151 3	0.10047 3	0.99517 2	0.98638 2
30	0.10860 3	0.10718 3	0.10590 3	0.10472 3	0.10365 3	0.10266 3	0.10176 3
31	0.11195 3	0.11048 3	0.10915 3	0.10794 3	0.10683 3	0.10581 3	0.10487 3
32	0.11530 3	0.11378 3	0.11241 3	0.11116 3	0.11001 3	0.10896 3	0.10798 3
33	0.11864 3	0.11708 3	0.11566 3	0.11437 3	0.11319 3	0.11210 3	0.11109 3
34	0.12199 3	0.12038 3	0.11892 3	0.11758 3	0.11636 3	0.11524 3	0.11420 3
35	0.12533 3	0.12367 3	0.12217 3	0.12079 3	0.11954 3	0.11838 3	0.11731 3
36	0.12868 3	0.12697 3	0.12542 3	0.12400 3	0.12271 3	0.12152 3	0.12042 3
37	0.13202 3	0.13026 3	0.12867 3	0.12721 3	0.12588 3	0.12466 3	0.12353 3
38	0.13536 3	0.13356 3	0.13192 3	0.13042 3	0.12905 3	0.12779 3	0.12663 3
39	0.13870 3	0.13685 3	0.13516 3	0.13363 2	0.13222 3	0.13093 3	0.12973 3
40	0.14204 3	0.14014 3	0.13841 3	0.13683 3	0.13539 3	0.13406 3	0.13284 3
41	0.14538 3	0.14343 3	0.14166 3	0.14004 3	0.13856 3	0.13719 3	0.13594 3
42	0.14872 3	0.14672 3	0.14490 3	0.14324 3	0.14172 3	0.14033 3	0.13904 3
43	0.15206 3	0.15001 3	0.14814 3	0.14645 3	0.14489 3	0.14346 3	0.14214 3
44	0.15539 3	0.15329 3	0.15139 3	0.14965 3	0.14806 3	0.14659 3	0.14524 3
45	0.15873 3	0.15658 3	0.15463 3	0.15285 3	0.15122 3	0.14972 3	0.14834 3
46	0.16206 3	0.15987 3	0.15787 3	0.15605 3	0.15438 3	0.15285 3	0.15143 3
47	0.16540 3	0.16315 3	0.16111 3	0.15925 3	0.15755 3	0.15598 3	0.15453 3
48	0.16873 3	0.16644 3	0.16435 3	0.16245 3	0.16071 3	0.15911 3	0.15763 3
49	0.17207 3	0.16972 3	0.16759 3	0.16565 3	0.16387 3	0.16223 3	0.16072 3
50	0.17540 3	0.17301 3	0.17083 3	0.16885 3	0.16703 3	0.16536 3	0.16382 3

(cont'd on next page)

TABLE 11-4 (cont'd)

5% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

n/m	30	31	32	33	34	35	36
2	0.11080 2	0.11022 2	0.10969 2	0.10919 2	0.10872 2	0.10828 2	0.10787 2
3	0.14846 2	0.14764 2	0.14687 2	0.14616 2	0.14550 2	0.14487 2	0.14429 2
4	0.18416 2	0.18310 2	0.18211 2	0.18119 2	0.18033 2	0.17952 2	0.17876 2
5	0.21872 2	0.21741 2	0.21620 2	0.21507 2	0.21401 2	0.21302 2	0.21209 2
6	0.25252 2	0.25097 2	0.24953 2	0.24818 2	0.24693 2	0.24576 2	0.24465 2
7	0.28578 2	0.28398 2	0.28231 2	0.28076 2	0.27930 2	0.27794 2	0.27667 2
8	0.31863 2	0.31659 2	0.31469 2	0.31292 2	0.31127 2	0.30972 2	0.30827 2
9	0.35118 2	0.34888 2	0.34675 2	0.34477 2	0.34291 2	0.34118 2	0.33955 2
10	0.38347 2	0.38093 2	0.37856 2	0.37636 2	0.37430 2	0.37237 2	0.37057 2
11	0.41556 2	0.41277 2	0.41017 2	0.40774 2	0.40548 2	0.40336 2	0.40137 2
12	0.44749 2	0.44443 2	0.44160 2	0.43895 2	0.43649 2	0.43417 2	0.43200 2
13	0.47927 2	0.47596 2	0.47289 2	0.47002 2	0.46734 2	0.46484 2	0.46249 2
14	0.51093 2	0.50737 2	0.50405 2	0.50096 2	0.49808 2	0.49537 2	0.49284 2
15	0.54249 2	0.53867 2	0.53511 2	0.53180 2	0.52870 2	0.52580 2	0.52308 2
16	0.57396 2	0.56987 2	0.56608 2	0.56254 2	0.55923 2	0.55614 2	0.55323 2
17	0.60535 2	0.60100 2	0.59696 2	0.59320 2	0.58968 2	0.58639 2	0.58329 2
18	0.63667 2	0.63206 2	0.62778 2	0.62379 2	0.62006 2	0.61656 2	0.61328 2
19	0.66793 2	0.66306 2	0.65853 2	0.65431 2	0.65037 2	0.64667 2	0.64320 2
20	0.69913 2	0.69400 2	0.68923 2	0.68478 2	0.68062 2	0.67672 2	0.67307 2
21	0.73029 2	0.72489 2	0.71987 2	0.71519 2	0.71082 2	0.70672 2	0.70288 2
22	0.76140 2	0.75574 2	0.75047 2	0.74556 2	0.74097 2	0.73667 2	0.73264 2
23	0.79248 2	0.78655 2	0.78103 2	0.77589 2	0.77108 2	0.76658 2	0.76235 2
24	0.82351 2	0.81732 2	0.81155 2	0.80618 2	0.80116 2	0.79645 2	0.79203 2
25	0.85452 2	0.84805 2	0.84204 2	0.83643 2	0.83119 2	0.82628 2	0.82167 2
26	0.88549 2	0.87876 2	0.87250 2	0.86666 2	0.86120 2	0.85608 2	0.85128 2
27	0.91644 2	0.90944 2	0.90293 2	0.89685 2	0.89117 2	0.88585 2	0.88086 2
28	0.94736 2	0.94009 2	0.93333 2	0.92702 2	0.92112 2	0.91560 2	0.91041 2
29	0.97826 2	0.97072 2	0.96371 2	0.95716 2	0.95105 2	0.94531 2	0.93993 2
30	0.10091 3	0.10013 3	0.99406 2	0.98728 2	0.98095 2	0.97501 2	0.96943 2
31	0.10400 3	0.10319 3	0.10244 3	0.10174 3	0.10108 3	0.10047 3	0.99891 2
32	0.10708 3	0.10625 3	0.10547 3	0.10475 3	0.10407 3	0.10343 3	0.10284 3
33	0.11017 3	0.10930 3	0.10850 3	0.10775 3	0.10705 3	0.10640 3	0.10578 3
34	0.11325 3	0.11236 3	0.11153 3	0.11076 3	0.11003 3	0.10936 3	0.10872 3
35	0.11633 3	0.11541 3	0.11456 3	0.11376 3	0.11302 3	0.11232 3	0.11166 3
36	0.11940 3	0.11846 3	0.11758 3	0.11676 3	0.11599 3	0.11528 3	0.11460 3
37	0.12248 3	0.12151 3	0.12060 3	0.11976 3	0.11897 3	0.11823 3	0.11754 3
38	0.12555 3	0.12456 3	0.12363 3	0.12276 3	0.12195 3	0.12119 3	0.12047 3
39	0.12863 3	0.12760 3	0.12665 3	0.12576 3	0.12492 3	0.12414 3	0.12341 3
40	0.13170 3	0.13065 3	0.12967 3	0.12875 3	0.12790 3	0.12710 3	0.12634 2
41	0.13477 3	0.13369 3	0.13269 3	0.13175 3	0.13087 3	0.13005 3	0.12927 3
42	0.13785 3	0.13674 3	0.13571 3	0.13474 3	0.13384 3	0.13300 3	0.13221 3
43	0.14092 3	0.13978 3	0.13872 3	0.13774 3	0.13681 3	0.13595 3	0.13514 3
44	0.14398 2	0.14282 3	0.14174 3	0.14073 3	0.13978 3	0.13890 3	0.13806 3
45	0.14705 3	0.14586 3	0.14476 3	0.14372 3	0.14275 3	0.14185 3	0.14099 3
46	0.15012 3	0.14890 3	0.14777 3	0.14671 3	0.14572 3	0.14479 3	0.14392 3
47	0.15319 3	0.15194 3	0.15078 3	0.14970 3	0.14869 3	0.14774 3	0.14685 3
48	0.15625 3	0.15498 3	0.15380 3	0.15269 3	0.15166 3	0.15068 3	0.14977 3
49	0.15932 3	0.15802 3	0.15681 3	0.15568 3	0.15462 3	0.15363 3	0.15270 3
50	0.16239 3	0.16106 3	0.15982 3	0.15867 3	0.15759 3	0.15657 3	0.15562 3

(cont'd on next page)

TABLE 11-4 (cont'd)

5% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

n / m	37	38	39	40	41	42	43
2	0.10748 2	0.10712 2	0.10677 2	0.10645 2	0.10614 2	0.10585 2	0.10557 2
3	0.14374 2	0.14322 2	0.14273 2	0.14227 2	0.14183 2	0.14142 2	0.14103 2
4	0.17805 2	0.17738 2	0.17675 2	0.17615 2	0.17559 2	0.17505 2	0.17455 2
5	0.21122 2	0.21040 2	0.20962 2	0.20889 2	0.20820 2	0.20754 2	0.20692 2
6	0.24362 2	0.24264 2	0.24172 2	0.24086 2	0.24003 2	0.23925 2	0.23852 2
7	0.27547 2	0.27434 2	0.27328 2	0.27227 2	0.27132 2	0.27041 2	0.26956 2
8	0.30690 2	0.30562 2	0.30441 2	0.30326 2	0.30218 2	0.30115 2	0.30018 2
9	0.33802 2	0.33657 2	0.33521 2	0.33393 2	0.33271 2	0.33156 2	0.33047 2
10	0.36887 2	0.36727 2	0.36576 2	0.36433 2	0.36298 2	0.36170 2	0.36049 2
11	0.39951 2	0.39775 2	0.39609 2	0.39452 2	0.39303 2	0.39163 2	0.39029 2
12	0.42997 2	0.42805 2	0.42623 2	0.42452 2	0.42290 2	0.42137 2	0.41991 2
13	0.46028 2	0.45819 2	0.45623 2	0.45437 2	0.45262 2	0.45095 2	0.44937 2
14	0.49046 2	0.48821 2	0.48609 2	0.48409 2	0.48220 2	0.48040 2	0.47870 2
15	0.52053 2	0.51812 2	0.51585 2	0.51370 2	0.51167 2	0.50974 2	0.50791 2
16	0.55050 2	0.54793 2	0.54550 2	0.54320 2	0.54103 2	0.53897 2	0.53702 2
17	0.58039 2	0.57765 2	0.57506 2	0.57262 2	0.57031 2	0.56812 2	0.56604 2
18	0.61020 2	0.60729 2	0.60455 2	0.60196 2	0.59951 2	0.59719 2	0.59498 2
19	0.63994 2	0.63687 2	0.63397 2	0.63123 2	0.62864 2	0.62618 2	0.62385 2
20	0.66963 2	0.66639 2	0.66333 2	0.66044 2	0.65771 2	0.65511 2	0.65265 2
21	0.69926 2	0.69585 2	0.69264 2	0.68960 2	0.68672 2	0.68399 2	0.68140 2
22	0.72884 2	0.72526 2	0.72189 2	0.71870 2	0.71568 2	0.71281 2	0.71009 2
23	0.75838 2	0.75463 2	0.75110 2	0.74776 2	0.74459 2	0.74159 2	0.73874 2
24	0.78788 2	0.78396 2	0.78026 2	0.77677 2	0.77346 2	0.77032 2	0.76734 2
25	0.81734 2	0.81325 2	0.80939 2	0.80575 2	0.80229 2	0.79902 2	0.79591 2
26	0.84676 2	0.84251 2	0.83849 2	0.83469 2	0.83109 2	0.82768 2	0.82443 2
27	0.87616 2	0.87173 2	0.86755 2	0.86360 2	0.85985 2	0.85630 2	0.85293 2
28	0.90553 2	0.90093 2	0.89658 2	0.89248 2	0.88859 2	0.88490 2	0.88139 2
29	0.93487 2	0.93009 2	0.92559 2	0.92133 2	0.91729 2	0.91346 2	0.90983 2
30	0.96418 2	0.95924 2	0.95457 2	0.95015 2	0.94597 2	0.94200 2	0.93823 2
31	0.99348 2	0.98836 2	0.98353 2	0.97896 2	0.97463 2	0.97052 2	0.96662 2
32	0.10228 3	0.10175 3	0.10125 3	0.10077 3	0.10033 3	0.99901 2	0.99498 2
33	0.10520 3	0.10465 3	0.10414 3	0.10365 3	0.10319 3	0.10275 3	0.10233 3
34	0.10812 3	0.10756 3	0.10703 3	0.10652 3	0.10605 3	0.10559 3	0.10516 3
35	0.11105 3	0.11046 3	0.10992 3	0.10940 3	0.10890 3	0.10844 3	0.10799 3
36	0.11397 3	0.11337 3	0.11280 3	0.11227 3	0.11176 3	0.11128 3	0.11082 3
37	0.11688 3	0.11627 3	0.11569 3	0.11514 3	0.11461 3	0.11412 3	0.11365 3
38	0.11980 3	0.11917 3	0.11857 3	0.11800 3	0.11747 3	0.11696 3	0.11647 3
39	0.12272 3	0.12207 3	0.12145 3	0.12087 3	0.12032 3	0.11979 3	0.11930 3
40	0.12563 3	0.12496 3	0.12433 3	0.12373 3	0.12317 3	0.12263 3	0.12212 3
41	0.12855 3	0.12786 3	0.12721 3	0.12660 3	0.12602 3	0.12546 3	0.12494 3
42	0.13146 3	0.13075 3	0.13009 3	0.12946 3	0.12886 3	0.12830 3	0.12776 3
43	0.13437 3	0.13365 3	0.13297 3	0.13232 3	0.13171 3	0.13113 3	0.13058 3
44	0.13728 2	0.13654 3	0.13584 3	0.13518 3	0.13455 3	0.13396 3	0.13340 3
45	0.14019 3	0.13943 3	0.13872 3	0.13804 3	0.13740 3	0.13679 3	0.13621 3
46	0.14310 3	0.14232 3	0.14159 3	0.14090 3	0.14024 3	0.13962 3	0.13903 3
47	0.14601 3	0.14521 3	0.14446 3	0.14376 3	0.14308 3	0.14245 3	0.14184 3
48	0.14891 3	0.14810 3	0.14734 3	0.14661 3	0.14593 3	0.14527 3	0.14465 3
49	0.15182 3	0.15099 3	0.15021 3	0.14947 3	0.14877 3	0.14810 3	0.14747 3
50	0.15472 3	0.15388 3	0.15308 3	0.15232 3	0.15161 3	0.15093 3	0.15028 3

(cont'd on next page)

TABLE 11-4 (cont'd)

5% POINTS FOR HOTELLING'S GENERALIZED T-SQUARE

n / m	44	45	46	47	48	49	50
2	0.10531 2	0.10506 2	0.10482 2	0.10459 2	0.10437 2	0.10416 2	0.10396 2
3	0.14065 2	0.14030 2	0.13996 2	0.13963 2	0.13932 2	0.13903 2	0.13875 2
4	0.17406 2	0.17360 2	0.17316 2	0.17275 2	0.17235 2	0.17197 2	0.17160 2
5	0.20632 2	0.20576 2	0.20522 2	0.20471 2	0.20422 2	0.20375 2	0.20331 2
6	0.23781 2	0.23714 2	0.23651 2	0.23590 2	0.23532 2	0.23476 2	0.23423 2
7	0.26874 2	0.26797 2	0.26723 2	0.26653 2	0.26586 2	0.26521 2	0.26460 2
8	0.29925 2	0.29837 2	0.29753 2	0.29673 2	0.29597 2	0.29524 2	0.29454 2
9	0.32943 2	0.32844 2	0.32750 2	0.32660 2	0.32574 2	0.32492 2	0.32414 2
10	0.35934 2	0.35824 2	0.35719 2	0.35620 2	0.35524 2	0.35433 2	0.35346 2
11	0.38902 2	0.38782 2	0.38667 2	0.38557 2	0.38452 2	0.38352 2	0.38256 2
12	0.41853 2	0.41721 2	0.41595 2	0.41475 2	0.41361 2	0.41252 2	0.41147 2
13	0.44787 2	0.44644 2	0.44508 2	0.44378 2	0.44254 2	0.44135 2	0.44022 2
14	0.47708 2	0.47554 2	0.47407 2	0.47267 2	0.47133 2	0.47005 2	0.46883 2
15	0.50617 2	0.50452 2	0.50294 2	0.50144 2	0.50000 2	0.49863 2	0.49732 2
16	0.53516 2	0.53339 2	0.53171 2	0.53010 2	0.52857 2	0.52710 2	0.52570 2
17	0.56406 2	0.56218 2	0.56039 2	0.55868 2	0.55704 2	0.55548 2	0.55398 2
18	0.59288 2	0.59088 2	0.58898 2	0.58717 2	0.58543 2	0.58377 2	0.58219 2
19	0.62163 2	0.61952 2	0.61750 2	0.61558 2	0.61375 2	0.61199 2	0.61032 2
20	0.65031 2	0.64808 2	0.64596 2	0.64393 2	0.64200 2	0.64015 2	0.63838 2
21	0.67893 2	0.67659 2	0.67435 2	0.67222 2	0.67019 2	0.66824 2	0.66637 2
22	0.70751 2	0.70504 2	0.70270 2	0.70046 2	0.69832 2	0.69627 2	0.69432 2
23	0.73603 2	0.73345 2	0.73099 2	0.72865 2	0.72641 2	0.72426 2	0.72221 2
24	0.76451 2	0.76181 2	0.75924 2	0.75679 2	0.75444 2	0.75220 2	0.75006 2
25	0.79295 2	0.79013 2	0.78745 2	0.78489 2	0.78244 2	0.78010 2	0.77786 2
26	0.82135 2	0.81842 2	0.81562 2	0.81295 2	0.81040 2	0.80796 2	0.80563 2
27	0.84972 2	0.84667 2	0.84376 2	0.84098 2	0.83833 2	0.83579 2	0.83336 2
28	0.87806 2	0.87489 2	0.87186 2	0.86897 2	0.86622 2	0.86358 2	0.86105 2
29	0.90637 2	0.90308 2	0.89994 2	0.89694 2	0.89408 2	0.89134 2	0.88872 2
30	0.93465 2	0.93124 2	0.92798 2	0.92488 2	0.92191 2	0.91907 2	0.91635 2
31	0.96291 2	0.95937 2	0.95601 2	0.95279 2	0.94972 2	0.94678 2	0.94396 2
32	0.99114 2	0.98749 2	0.98400 2	0.98068 2	0.97750 2	0.97446 2	0.97155 2
33	0.10194 3	0.10156 3	0.10120 3	0.10085 3	0.10053 3	0.10021 3	0.99911 2
34	0.10475 3	0.10436 3	0.10399 3	0.10364 3	0.10330 3	0.10298 3	0.10266 3
35	0.10757 3	0.10717 3	0.10679 3	0.10642 3	0.10607 3	0.10574 3	0.10542 3
36	0.11039 3	0.10997 3	0.10958 3	0.10920 3	0.10884 3	0.10850 3	0.10817 3
37	0.11320 3	0.11277 3	0.11237 3	0.11198 3	0.11161 3	0.11125 3	0.11091 3
38	0.11601 3	0.11557 3	0.11516 3	0.11476 3	0.11438 3	0.11401 3	0.11366 3
39	0.11882 3	0.11837 3	0.11794 3	0.11753 3	0.11714 3	0.11677 3	0.11641 3
40	0.12163 3	0.12117 3	0.12073 3	0.12031 3	0.11990 3	0.11952 3	0.11915 3
41	0.12444 3	0.12397 3	0.12351 3	0.12308 3	0.12267 3	0.12227 3	0.12189 3
42	0.12725 3	0.12676 3	0.12630 3	0.12585 3	0.12543 3	0.12502 3	0.12463 3
43	0.13005 3	0.12955 3	0.12908 3	0.12862 3	0.12819 3	0.12777 3	0.12737 3
44	0.13286 2	0.13235 3	0.13186 3	0.13139 3	0.13094 3	0.13052 3	0.13011 3
45	0.13566 3	0.13514 3	0.13464 3	0.13416 3	0.13370 3	0.13326 3	0.13285 3
46	0.13846 3	0.13793 3	0.13741 3	0.13692 3	0.13646 3	0.13601 3	0.13558 3
47	0.14126 3	0.14072 3	0.14019 3	0.13969 3	0.13921 3	0.13875 3	0.13832 3
48	0.14406 3	0.14350 3	0.14297 3	0.14246 3	0.14197 3	0.14150 3	0.14105 3
49	0.14686 3	0.14629 3	0.14574 3	0.14522 3	0.14472 3	0.14424 3	0.14378 3
50	0.14966 3	0.14908 3	0.14852 3	0.14798 3	0.14747 3	0.14698 3	0.14651 3

TABLE 11-5

TABLE OF COEFFICIENTS FOR APPROXIMATING 1% AND 5% UPPER PROBABILITY LEVELS FOR HOTELLING'S GENERALIZED T^2 STATISTICS

(Bivariate Case With $m > 50$)

COEFFICIENTS FOR QUADRATIC FORMULA—1% LEVEL

$$T^2 \approx am^2 + bm + c$$
$$51 \leq m \leq 101$$

n	a	b	c
2	.1128472 -1	.3141341 5	-.1553434 5
3	.0000000 0	.2999800 3	-.2989990 1
4	-.3385461 -4	.6308968 2	.2698526 2
5	-.3645817 -4	.2884334 2	.2323330 2
6	-.3802073 -4	.1800412 2	.2035420 2
7	-.4123275 -4	.1313480 2	.1855297 2
8	-.4644112 -4	.1047653 2	.1739104 2
9	-.4999992 -4	.8835404 1	.1662274 2
10	-.5468768 -4	.7733995 1	.1609091 2
11	-.5972230 -4	.6948779 1	.1571550 2
12	-.6467011 -4	.6362953 1	.1544823 2
13	-.6987837 -4	.5910259 1	.1525495 2
14	-.7517356 -4	.5550468 1	.1511607 2
15	-.8072911 -4	.5257943 1	.1501600 2
16	-.8611112 -4	.5015521 1	.1494671 2
17	-.9157979 -4	.4811431 1	.1490013 2
18	-.9704868 -4	.4637270 1	.1487106 2
19	-.1025173 -3	.4486913 1	.1485548 2
20	-.1083334 -3	.4355842 1	.1484875 2
21	-.1137153 -3	.4240466 1	.1485263 2
22	-.1192708 -3	.4138186 1	.1486211 2
23	-.1248265 -3	.4046870 1	.1487661 2
24	-.1302951 -3	.3964821 1	.1489561 2
25	-.1357639 -3	.3890702 1	.1491762 2
26	-.1412326 -3	.3823408 1	.1494215 2
27	-.1466146 -3	.3762023 1	.1496865 2
28	-.1519097 -3	.3705791 1	.1499710 2
29	-.1572049 -3	.3654095 1	.1502653 2
30	-.1624132 -3	.3606393 1	.1505708 2
31	-.1677083 -3	.3562260 1	.1508743 2
32	-.1730035 -3	.3521298 1	.1511796 2
33	-.1780382 -3	.3483137 1	.1514979 2
34	-.1832466 -3	.3447552 1	.1518079 2
35	-.1882812 -3	.3414244 1	.1521236 2
36	-.1934028 -3	.3383040 1	.1524292 2
37	-.1982639 -3	.3353690 1	.1527462 2
38	-.2031250 -3	.3326069 1	.1530562 2
39	-.2079861 -3	.3300027 1	.1533619 2
40	-.2128472 -3	.3275435 1	.1536601 2
41	-.2175347 -3	.3252147 1	.1539622 2
42	-.2223958 -3	.3230109 1	.1542487 2
43	-.2271702 -3	.3209194 1	.1545306 2
44	-.2315972 -3	.3189273 1	.1548247 2
45	-.2361979 -3	.3170348 1	.1551028 2
46	-.2406250 -3	.3152298 1	.1553847 2
47	-.2452257 -3	.3135115 1	.1556486 2
48	-.2496528 -3	.3118690 1	.1559138 2
49	-.2539931 -3	.3102982 1	.1561764 2
50	-.2583334 -3	.3087958 1	.1564325 2
51	-.2625868 -3	.3073561 1	.1566868 2
52	-.2668403 -3	.3059768 1	.1569320 2
53	-.2710937 -3	.3046537 1	.1571723 2
54	-.2753472 -3	.3033840 1	.1574036 2
55	-.2794271 -3	.3021616 1	.1576377 2
56	-.2834201 -3	.3009853 1	.1578699 2
57	-.2875001 -3	.2998550 1	.1580894 2
58	-.2914930 -3	.2987653 1	.1583086 2
59	-.2953993 -3	.2977141 1	.1585264 2
60	-.2993056 -3	.2967008 1	.1587377 2
61	-.3032986 -3	.2957249 1	.1589368 2
62	-.3070313 -3	.2947790 1	.1591439 2
63	-.3106771 -3	.2938641 1	.1593497 2
64	-.3143229 -3	.2929801 1	.1595493 2
65	-.3179688 -3	.2921252 1	.1597441 2
66	-.3215277 -3	.2912971 1	.1599358 2
67	-.3251736 -3	.2904968 1	.1601182 2
68	-.3288195 -3	.2897219 1	.1602950 2
69	-.3323785 -3	.2889701 1	.1604699 2
70	-.3359375 -3	.2882411 1	.1606419 2
71	-.3394098 -3	.2875332 1	.1608116 2
72	-.3427083 -3	.2868440 1	.1609857 2
73	-.3460937 -3	.2861762 1	.1611503 2
74	-.3493924 -3	.2855269 1	.1613124 2
75	-.3527778 -3	.2848975 1	.1614655 2
76	-.3559896 -3	.2842830 1	.1616261 2
77	-.3593750 -3	.2836885 1	.1617708 2
78	-.3625000 -3	.2831067 1	.1619256 2
79	-.3657119 -3	.2825421 1	.1620693 2
80	-.3687500 -3	.2819900 1	.1622202 2
81	-.3718750 -3	.2814540 1	.1623613 2
82	-.3749132 -3	.2809310 1	.1625014 2
83	-.3779514 -3	.2804214 1	.1626386 2
84	-.3809028 -3	.2799235 1	.1627772 2
85	-.3839410 -3	.2794397 1	.1629066 2
86	-.3868924 -3	.2789669 1	.1630367 2
87	-.3897569 -3	.2785047 1	.1631657 2
88	-.3926216 -3	.2780537 1	.1632922 2
89	-.3954861 -3	.2776135 1	.1634165 2
90	-.3983507 -3	.2771842 1	.1635356 2
91	-.4012153 -3	.2767649 1	.1636526 2
92	-.4039930 -3	.2763545 1	.1637690 2
93	-.4066840 -3	.2759521 1	.1638880 2
94	-.4096354 -3	.2755639 1	.1639827 2
95	-.4124132 -3	.2751814 1	.1640887 2
96	-.4149306 -3	.2748031 1	.1642084 2
97	-.4175347 -3	.2744351 1	.1643170 2
98	-.4199653 -3	.2740724 1	.1644361 2
99	-.4224826 -3	.2737191 1	.1645443 2
100	-.4250868 -3	.2733748 1	.1646448 2

(cont'd on next page)

TABLE 11-5 (cont'd)

COEFFICIENTS FOR QUADRATIC FORMULA—5% LEVEL

$$T^2 \approx am^2 + bm + c$$
$$51 \leq m \leq 101$$

n	a	b	c
2	.3472169 -3	.1256467 4	-.6237031 3
3	.1735985 -5	.5990428 2	-.2988857 1
4	-.1128473 -4	.2147326 2	.6552336 1
5	-.1527762 -4	.1278487 2	.7528524 1
6	-.1892375 -4	.9334072 1	.7680655 1
7	-.2239593 -4	.7550043 1	.7714972 1
8	-.2595493 -4	.6477283 1	.7740685 1
9	-.2968756 -4	.5766432 1	.7775570 1
10	-.3350692 -4	.5262601 1	.7821199 1
11	-.3741314 -4	.4887514 1	.7874495 1
12	-.4123265 -4	.4597616 1	.7934323 1
13	-.4513893 -4	.4366925 1	.7997431 1
14	-.4913192 -4	.4178982 1	.8062396 1
15	-.5303825 -4	.4022866 1	.8128778 1
16	-.5711812 -4	.3891126 1	.8194236 1
17	-.6119793 -4	.3778415 1	.8259405 1
18	-.6519103 -4	.3680847 1	.8324144 1
19	-.6935761 -4	.3595572 1	.8386907 1
20	-.7326389 -4	.3520331 1	.8449366 1
21	-.7725688 -4	.3453468 1	.8510192 1
22	-.8133685 -4	.3393653 1	.8568674 1
23	-.8532995 -4	.3339789 1	.8626001 1
24	-.8914930 -4	.3291008 1	.8682479 1
25	-.9314235 -4	.3246657 1	.8736368 1
26	-.9704862 -4	.3206120 1	.8789114 1
27	-.1009549 -3	.3168933 1	.8840010 1
28	-.1047743 -3	.3134681 1	.8889501 1
29	-.1085938 -3	.3103029 1	.8937791 1
30	-.1124132 -3	.3073697 1	.8984219 1
31	-.1162326 -3	.3046433 1	.9029148 1
32	-.1198784 -3	.3021001 1	.9073377 1
33	-.1235243 -3	.2997239 1	.9116095 1
34	-.1273438 -3	.2975012 1	.9156710 1
35	-.1309028 -3	.2954119 1	.9197122 1
36	-.1343750 -3	.2934460 1	.9236528 1
37	-.1381077 -3	.2915977 1	.9273513 1
38	-.1414063 -3	.2898471 1	.9311359 1
39	-.1449653 -3	.2881961 1	.9346621 1
40	-.1483507 -3	.2866305 1	.9381621 1
41	-.1519097 -3	.2851486 1	.9414508 1
42	-.1551215 -3	.2837366 1	.9448098 1
43	-.1585070 -3	.2823968 1	.9479624 1
44	-.1618056 -3	.2811200 1	.9510656 1
45	-.1651910 -3	.2799047 1	.9540045 1
46	-.1684028 -3	.2787427 1	.9569434 1
47	-.1715278 -3	.2776313 1	.9598506 1
48	-.1747396 -3	.2765701 1	.9626071 1
49	-.1777778 -3	.2755521 1	.9653538 1
50	-.1809028 -3	.2745781 1	.9679684 1

n	a	b	c
51	-.1839410 -3	.2736431 1	.9705463 1
52	-.1870660 -3	.2727470 1	.9730073 1
53	-.1899305 -3	.2718827 1	.9755228 1
54	-.1929687 -3	.2710547 1	.9778595 1
55	-.1958333 -3	.2702558 1	.9802088 1
56	-.1987847 -3	.2694880 1	.9824548 1
57	-.2017361 -3	.2687485 1	.9846259 1
58	-.2045139 -3	.2680335 1	.9868034 1
59	-.2072049 -3	.2673424 1	.9889691 1
60	-.2100695 -3	.2666781 1	.9909847 1
61	-.2128472 -3	.2660356 1	.9929647 1
62	-.2155382 -3	.2654137 1	.9949229 1
63	-.2181423 -3	.2648111 1	.9968731 1
64	-.2208333 -3	.2642296 1	.9987200 1
65	-.2235243 -3	.2636668 1	.1000521 2
66	-.2261284 -3	.2631209 1	.1002301 2
67	-.2285591 -3	.2625894 1	.1004117 2
68	-.2311632 -3	.2620768 1	.1005808 2
69	-.2335938 -3	.2615774 1	.1007518 2
70	-.2361111 -3	.2610942 1	.1009160 2
71	-.2387153 -3	.2606270 1	.1010664 2
72	-.2411459 -3	.2601705 1	.1012245 2
73	-.2435764 -3	.2597274 1	.1013777 2
74	-.2459201 -3	.2592957 1	.1015304 2
75	-.2482639 -3	.2588760 1	.1016785 2
76	-.2505208 -3	.2584670 1	.1018265 2
77	-.2526910 -3	.2580677 1	.1019754 2
78	-.2551215 -3	.2576833 1	.1021050 2
79	-.2573785 -3	.2573063 1	.1022423 2
80	-.2596354 -3	.2569393 1	.1023756 2
81	-.2618924 -3	.2565823 1	.1025019 2
82	-.2640625 -3	.2562330 1	.1026319 2
83	-.2659723 -3	.2558888 1	.1027713 2
84	-.2682292 -3	.2555584 1	.1028876 2
85	-.2703125 -3	.2552339 1	.1030102 2
86	-.2723090 -3	.2549161 1	.1031337 2
87	-.2743924 -3	.2546073 1	.1032485 2
88	-.2764757 -3	.2543061 1	.1033611 2
89	-.2786459 -3	.2540134 1	.1034650 2
90	-.2806424 -3	.2537248 1	.1035778 2
91	-.2825521 -3	.2534418 1	.1036899 2
92	-.2845486 -3	.2531666 1	.1037956 2
93	-.2864584 -3	.2528965 1	.1039018 2
94	-.2882813 -3	.2526311 1	.1040116 2
95	-.2905382 -3	.2523787 1	.1040905 2
96	-.2924479 -3	.2521261 1	.1041915 2
97	-.2944445 -3	.2518808 1	.1042812 2
98	-.2961806 -3	.2516369 1	.1043866 2
99	-.2979166 -3	.2513983 1	.1044863 2
100	-.3000000 -3	.2511700 1	.1045660 2

Example 11-3:

A new artillery projectile was designed to replace the old standard one. The new projectile was made longer so that it would have more explosive, the surface finish was much smoother, and it was to be fired from guns with a higher twist of rifling to give improved stability. Unfortunately, only a few projectiles of each of the old and new types were available for test in this particular part of the overall program. Thus 10 of the old, or standard, projectiles were fired to find range and deflection deviations along with only eight projectiles of the proposed artillery rounds. The results of the firing program are given in Table 11-6. It was expected that the newly designed artillery projectiles should give a smaller dispersion in range and deflection and that they should also give increased ranges due to their improved stability and surface finish. Does there exist, therefore, any substantial evidence to support these hypotheses?

The various questions arising here may be answered easily by carrying out an analysis using the Hotelling Generalized T^2 statistic and the Hotelling Multivariate Studentized t or T_M^2 for mean values.

TABLE 11-6

RANGE AND DEFLECTION IMPACT POSITIONS FOR NEW AND OLD ARTILLERY PROJECTILES

Standard Projectile ("Old")		Proposed Projectile ("New")	
Range x_{1p}, m	Deflection x_{2p}, m	Range x'_{1p}, m	Deflection x'_{2p}, m
6351	2	6457	20
6331	7	6494	12
6355	6	6482	14
6319	0	6447	22
6242	0	6382	7
6323	6	6430	15
6246	10	6381	12
6294	-5	6348	11
6354	11		
6283	5		

The pertinent calculations based on Table 11-6 are

Old Sample		New Sample	
$N = 10$		$M = 8$	
$\bar{x}_1 = 6309.800$	$\bar{x}_2 = 4.200$	$\bar{x}'_1 = 6427.625$	$\bar{x}'_2 = 14.125$.

We next estimate the variance-covariance matrix of the old population, i.e., $[\sigma_{ij}]$, by using the old bivariate sample to obtain

$$[s_{ij}] = \begin{bmatrix} 1781.95556 & 41.60000 \\ 41.60000 & 24.40000 \end{bmatrix}.$$

The inverse of this matrix gives the v_{ij} matrix, which is

$$[v_{ij}] = [s_{ij}]^{-1} = \begin{bmatrix} 0.00058444 & -0.00099643 \\ -0.00099643 & 0.04268243 \end{bmatrix}$$

The variance-covariance matrix for the new artillery projectile, or new sample, is from Eq. 11-37

$$[s'_{ij}] = \begin{bmatrix} 2743.12500 & 121.76786 \\ 121.76786 & 23.83929 \end{bmatrix}$$

which is used to calculate T_D^2, whereas we calculate the overall T_0^2 by using Eq. 11-44 to obtain the variance-covariance matrix of the z_{ip}, z_{jp}, which is

$$[s''_{ij}] = \begin{bmatrix} 16282.965 & 1275.960 \\ 1275.960 & 119.365 \end{bmatrix}$$

We are now ready to calculate T_0^2 from Eq. 11-45, and it is easily seen that eight times the trace of the product of $[v_{ij}]$ and $[s''_{ij}]$ is simply

$$\begin{aligned} T_0^2 &= 8[(0.00058444)\,(16282.965) + (-0.00099643)\,(1275.96) \\ &\quad (-0.00099643)\,(1275.960) + (0.04268243)\,(119.365)] \\ &= 8(8.2444 + 3.8234) = 96.547. \end{aligned}$$

This is a test of whether the new projectiles and the old, or standard ones, are equivalent in all respects, i.e., have the same variances and covariance, and their population mean ranges are equal and the deflection shifts are equal. We refer the observed value 96.547 to Table 11-4—the percentage points of Hotelling's Generalized T^2 with $m = 8$ for the total new sample size, and $n = 9$ for the df for the old sample variances—and obtain a value of $T_0^2(5\%) = 53.173$ for the 5% level. Moreover, the 1% level is only $T_0^2(1\%) = 86.066$. We therefore decide to reject the null, or tested, hypothesis and make the judgment that the standard and proposed projectiles are not equivalent in either their mean values of their variances and covariance, or perhaps even both. Therefore, we must analyze the data further.

As a matter of some interest at this point, we might use the quadratic computation of Eq. 11-43 and Table 11-5 to see just how far off our predicted percentage point is for these very small sample sizes. We have in this connection that at the 1% level

$$T_0^2(1\%) \approx (-0.00005)(64) + (8.835)(8) + 16.623 = 87.3,$$

which is only 1.4 higher than the exact value! Better accuracy could be expected for higher values of m and n.

Our problem now is to determine whether the dispersion parameters of the proposed and standard projectiles are different or whether their mean range and mean deflection values are unequal, or perhaps both of these, or finally whether we may have a small, chance variation or accidental occurrence. Before a test of mean values we should establish whether or not the old and new sample covariance matrices are equivalent. This is done by calculating the value of T_D^2 from

$$T_D^2 = (M-1)\sum_{i=1}^{2}\sum_{j=1}^{2} v_{ij}s'_{ij} = m\mathrm{tr}\{[v_{ij}][s'_{ij}]\} \qquad (11\text{-}46)$$

which is equivalent to Eq. 11-31 or Eq. 11-38. Thus the calculation of the observed T_D^2 is found to be

$$\begin{aligned} T_D^2 &= 7\mathrm{tr}\begin{bmatrix} 0.00058444 & -0.0009643 \\ -0.00099643 & 0.0426824 \end{bmatrix}\begin{bmatrix} 2743.1250 & 121.7679 \\ 121.7679 & 23.8393 \end{bmatrix} \\ &= 7(1.486 + 0.896) = 16.67. \end{aligned}$$

Referring to Table 11-4, we find for $m = 7$ and $n = 9$ that the 1% probability level of Hotelling's T_D^2 is 77.030, and the 5% level is 47.289, so that we conclude that the two samples originate from normal bivariate populations with identical dispersion, or variance-covariance, matrices. Hence the new artillery rounds do not give smaller dispersion in either range or deflection. Having established this, we then proceed to examine whether there is a difference in the average ranges or average deflections of the standard and proposed projectiles. As a matter of fact, it is noted that the new round gives a somewhat longer range (6428 — 6310 = 118 m), and the proposed projectiles may deflect farther to the right. To test for equal C of I's, we use Hotelling's Studentized T^2, or T_S^2, and initially, for illustration, the variance-covariance matrix of the old sample based on $n = N - 1 = 9$ df. Since, however, the centroid for the new sample is a single point, the M for the new sample is taken appropriately as unity, i.e., $M = 1$. Then, from Eq. 11-24, we have

$$T_S^2 = \frac{NM}{N+M}[\bar{z}]^T [v_{ij}(\text{old})]\,[\bar{z}] \tag{11-47}$$

$$= \frac{(10)\,(1)}{10+1}\,[117.825 \quad 9.925]\begin{bmatrix} 0.00058444 & -0.0009643 \\ -0.0009643 & 0.0426824 \end{bmatrix}\begin{bmatrix} 117.825 \\ 9.925 \end{bmatrix}$$

$$= 9.148.$$

Hence, from Eq. 11-22

$$F(2,8) = \frac{8}{(2)\,(9)}\,(9.148) = 4.066.$$

Now the 5% point for $F(2,8)$ is

$$\text{upper } F_{0.05}(2,8) = 4.46$$

so that with the use of only the old sample to estimate the variance-covariance matrix, we are not able to detect any difference between the means in range and means in deflection. However, since we established that the old, or standard, and the proposed projectiles have equivalent covariance and variances in range and in deflection, then we should pool (add) the SS of the two samples in order to base the variance-covariance matrix on the entire number of df available, i.e.,

$$N - 1 + M - 1 = 16 \text{ df.}$$

The new $[s_{ij}]$ is

$$[s_{ij}] = \begin{bmatrix} 2202.4672 & 76.6734 \\ 76.6734 & 24.1547 \end{bmatrix}$$

and

$$T_S^2 = \frac{NM}{N+M}\,[\bar{z}]^T [v_{ij}\,(\text{new})]\,[\bar{z}]$$

$$= \frac{10(8)}{18}\,[117.825 \quad 9.925]\begin{bmatrix} 0.0005104 & -0.0016203 \\ -0.0016203 & 0.0465430 \end{bmatrix}\begin{bmatrix} 117.825 \\ 9.925 \end{bmatrix}$$

$$= \frac{80}{18}\,(7.881) = 35.026.$$

Finally, transforming T_S^2 to an equivalent Snedecor F, we obtain

$$F(2,15) = (N + M - 3)T_S^2/[2(N + M - 2)] = 15(35.029)/32 = 16.42$$

but the upper 1% probability level of F is

$$F_{0.01}(2,15) = 6.36$$

so we must conclude that the proposed artillery projectiles have significantly longer ranges (by 118 m) and deflect more to the right (by about 14 m). In this connection, it would be well to check both of these conclusions by using separate Student's t tests for the ranges and deflections, ignoring any correlation. If this were done, it would be found that the average range of the proposed projectiles is significantly greater than that of the old projectiles and also that the deflections differ as indicated.

As a final point of interest, we found, when using the Hotelling T_D^2 test, that the variance-covariance matrices of the old and new samples are equivalent in all respects. However, suppose we had found that the dispersion matrices were significantly different. Then, somewhat of a problem would arise, and we would have to decide whether to use the old sample alone or the new sample results alone to conduct our T_M^2 significance test. That is to say for this example that we would not have been able to detect any differences in either range or deflection—unless correlation could have been ignored and we used Student's t tests separately. Nevertheless, it is conjectured that had we used the T_M^2 test anyway, ignoring significantly different covariances based on the T_D^2 test, the resulting procedure would have been very robust and, hence, rather dependable.

11-4 SUMMARY

In this chapter we have described Wilks' statistics for testing the equality of population means, the equality of variances, and the equality of covariances for single multivariate normal distributions. These tests are needed to judge the dispersion values and levels of the characteristics of a bivariate or multivariate normal sample, especially for the case of suspected correlation or dependence between the characteristics. An example has been given to analyze the jump of the first and second bullets in rapid fire from an M16 rifle.

There are many applications of Army interest for which it becomes necessary to make an overall comparison of two different multivariate normal samples. For the comparison of true means of corresponding characteristics of bivariate or multivariate normal samples, Hotelling's Multivariate Studentized statistic (T_S^2) is used under the assumption that the variance-covariance matrices of the two samples are equivalent, or nearly so. Hotelling's T_S^2 is especially required when there is some correlation between two or more of the characteristics, and yet it can be transformed to the Snedecor F-type statistic or test.

For a comparison of the two dispersion matrices or the variance-covariance matrices of the two sampled populations, another statistic, known as Hotelling's Generalized T_D^2, is required. Then, for an overall or combined test for both the equality of variance-covariance matrices *and* the equality of corresponding true means of the individual characteristics, the proper significance test involves a quantity we have defined as Hotelling's Generalized T_0^2(total) statistic. In fact, we have that $T_0^2 = T_D^2 + T_M^2$, where T_M^2 is a statistic for testing the equality of the corresponding characteristic means and is directly relatable to Hotelling's T_S^2 sample statistic. New tables of percentage points are necessary for the T_D^2 and T_0^2 Hotelling statistics, which depend on Karl Pearson's incomplete beta function ratio. In spite of this complication and the fact that T_0^2 and T_D^2 depend on the number of df n in the old sample and m for the new sample, it has been found that for fixed n the percentage points are very nearly linear as a function of the T^2's for different m! Consequently, the size of the tables of probability levels for m, n greater than 50 can be reduced considerably by providing a short table of coefficients. A very extensive and highly informative example is given that covers the analysis of dispersion patterns and ranges to ground impact of some standard and some newly proposed artillery projectiles.

In Army statistical work there should be many diverse types of applications of the multivariate statistical theory presented in this chapter.

REFERENCES

1. DARCOM-P 706-101, Engineering Design Handbook, *Army Weapon Systems Anaylsis, Part One*, November 1977.
2. DARCOM-P 706-102, Engineering Design Handbook, *Army Weapon Systems Analysis, Part Two*, October 1979.
3. Frank E. Grubbs, *Statistical Measures of Accuracy for Riflemen and Missile Engineers*, Edwards Brothers, Ann Arbor, MI, November 1964.
4. Samuel S. Wilks, "Sample Criteria for Testing Equality of Means, Equality of Variances, and Equality of Covariances in a Normal Multivariate Distribution", The Annals of Mathematical Statistics **17**, 257-81 (September 1946).
5. Harold Hotelling, "Multivariate Quality Control, Illustrated by the Air Testing of Sample Bombsights", Chapter 3 of *Selected Techniques of Statistical Analysis*, Editors: C. Eisenhart, M. W. Hastay, and W. A. Wallis, McGraw-Hill Book Company, Inc., New York, NY, 1947.
6. Harold Hotelling, "The Generalization of Student's Ratio", The Annals of Mathematical Statistics **2**, 360-78 (1931).
7. Harold Hotelling, "A Generalized T^2 Test and Measure of Multivariate Dispersion", *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, CA, 1951, 23-41.
8. J. Stuart Hunter, *Hotelling's T^2 Statistics*, Unpublished Manuscript, US Army Ballistic Research Laboratories, Aberdeen Proving Ground, MD, 1952.
9. Frank E. Grubbs, *Tables of 1% and 5% Probability Levels for Hotelling's Generalized T^2 Statistic (Bi-Variate Case)*, BRL Technical Note No. 926, US Army Ballistic Research Laboratories, Aberdeen Proving Ground, MD, July 1954.
10. Frank E. Grubbs, Helen J. Coon, J. Stuart Hunter, and H. K. Crowder, *Tables of 1% and 5% Probability Levels for Hotelling's Generalized T^2 Statistic (Bivariate Case)*, Unpublished Manuscript, US Army Ballistic Research Laboratories, Aberdeen, MD, 1968.
11. John Wishart, "The Generalized Product Moment Distribution in Samples from a Normal Multivariate Population", Biometrika **20A**, 32-52 (1928).
12. Karl Pearson, Editor, *Tables of the Incomplete Beta Function*, Cambridge University Press, Cambridge, England, 1934.

INDEX

Accuracy of measurement, 2-5
Accuracy test, 2-25
Analysis of variance (ANOVA) example, 4-54
Approximate chi-square distribution, 4-25
Bartlett-Kendall test for variances, 4-36
Bartlett's test, 4-33
Behrens-Fisher problem, 4-45
Berkson's controlled independent variable, 6-15
Best linear unbiased estimates (BLUE), 7-9
    Table of coefficients, 7-15
Bias changes, instrumental, 6A-6
Biases
    of estimators, 4-12
    of measurements, 2-6, 6A-2
Biases, varying, 6A-5
Binomial distribution, 4-17
Binomial population comparisons, 5-12
Bivariate and multivariate random variables, 11-3
Bounds, mathematical
    for range to standard deviation, 3-5
    for residuals, 3-6
Bradley-Brindley analysis, three-instrument imprecisions, 2-50
Britt-Luecke algorithm for generalized least squares, 6-53
Cadwell's test, 4-34
Calibration hierarchy, 2-46
Calibration of instruments, 2-46
Chi-square distribution, 4-15
Chi-square percentage points, 4-18, 4-19
Cochran's test of variances, 4-34
Combining independent tests, 4-56
Comparison of binomial populations, 5-12
Comparison of measuring instruments, 2-5, 2-32
Comparison of normal populations, 4-5
Complete rundown test, 9-7
Confidence bounds
    on imprecision and inaccuracy parameters, 2-34
    on normal population means, 4-41
    on normal population variances, 4-18
    on reliability, 8-14
Contingency tables, 5-3, 5-9
Controlled independent variable, 6-15
Criteria for outliers, 3-11
Description of handbook chapters, 1-1
Design and analysis of experiments, 4-49
Determination of sample size, 8-3
    Basic principles, 8-6
    Binomial and Poisson populations, 8-8
    Variance estimation for normal populations, 8-18
    Analysis of variance tests, 8-36
    Exponential populations, 8-41
Distribution of largest sample value, 7-5
Distribution of range, 7-6
Distribution of rth order statistic, 7-6
Distribution of smallest sample value, 7-5
Dixon's outlier criteria, 3-16, 3-17
Duncan's multiple range test, 7-10
Double dichotomy, 5-31
Efficiency of estimators, 4-12
Equal spacing of independent variables, 6-13
Error in both independent and dependent variables, 6-14, 6-23
Errors of measurement, 2-5, 2-6
Estimation, 4-5
F distribution (Snedecor-Fisher), 4-26
F percentage points, 4-29
F test (Snedecor), 6-9
Fisher exact test, contingency tables, 5-10
Fitting a line, 6-5
Fitting a parabola, 6-29
Fitting a plane, 6-25, 6-32
Functional relations, 6-4, 6-14, 10-1
General linear model, regression, 6-45
General two-way contingency tables, 5-38
Generalized least squares, 6-50
Grubbs' outlier tests, 3-12
Gumbel's extreme value distribution, 7-23
Hartley's test of variances, 4-34
Hawkins' tests, 3-28
Higher order contingency tables, 5-43
Homoscedasticity, 4-38
Hotelling's generalized T^2 percentage points, 11-17
Hotelling's generalized T^2 statistics, (T_0^2, T_D^2, T_M^2), 11-13
    Linear relationship, 11-15, 11-16
    Quadratic approximation, 11-16
Hotelling's multivariate studentized t statistics (T_S^2), 11-10
Imprecision, 2-1
Imprecision test, 2-27, 2-30
Inaccuracy, 2-1
Independence and interaction, contingency tables, 5-33
Independent standard deviation, 3-40
Interlaboratory testing (round-robin testing), 2-43
Introduction to handbook, 1-1
Karni-Weissman analysis, 6-24
Kullback's minimum discrimination information statistics, 5-34, 5-40
Kurtosis, 3-37
Langlie one-shot test strategy, 9-8
Largest observation, distribution of, 7-5
Largest observation, test of, 3-12, 3-17
Least squares, 6-1

INDEX (cont'd)

Least squares, both variables subject to error, 6-21
Least squares applied to precision and accuracy of ozone measurements, 6A-1
Least squares vs physical modeling, 10-1
Lieblein's ratio, 3-9
Limit velocity analysis, 10-1
Line fitting, 6-5
Linear estimation, 7-9
Logistic distribution, 9-18
Loglinear analysis, 5-43
Many outliers test, 3-26
Maximum likelihood estimation, (examples), 9-14
    Normal distribution, 9-15
    Logistic distribution, 9-18
    Weibull distribution, 9-22
McDonald, Davis, and Milliken tables, 5-21
Mean deviation, 4-8
Mean square error, 4-13
Model building, 10-1
Moment properties, 4-14
Moments of exponential, gamma, and Weibull populations, 7-11
Multi-instrument case, 2-40
Multiple regression, 6-45
    General linear model, 6-45
Multiple significance tests, 4-57
Multivariate statistical analyses, 11-1
Negative variance estimators, 2-15
Nonlinear regression, 6-50, 6-53
Operating characteristic curves (power curves), 8-22, 8-23, 8-26, 8-30, 8-32, 8-39
Order statistics, 7-3
Order statistics and reliability, 7-34
Orthogonal polynomial examples, 6-36, 6A-9
Orthogonal polynomials, 6-34
    Tables of, 6-37
Outlier bounds, 3-5
Outlying observations, 3-1
Ozone distribution analysis, 6A-2
Ozone rocket sonde intercomparison, 6A-2
Parabola (quadratic) fitting, 6-29
Plane fitting, 6-25
Poisson confidence bounds, 5-8
Poisson-chi-square relationship, 4-17
Power curves *See:* Operating characteristics
Power of 2 × 2 contingency tables, 5-37
Precision of measurements, 2-5
Probability plots, 3-54
Product variability, 2-11
Quantal response data analysis *See also:* Sensitivity analysis, 9-3
    Computer programs, 9A-1
Quasi-range, 7-7
Radial order statistics, 7-35
Range, maximum dispersion, 4-10, 7-4, 7-7
Range test, 3-18
Regression, 6-4, 6-5
Robbins-Monro stochastic technique, 9-9
Role of the statistician, 10-1
Rosner's tests (outliers), 3-28
Sample size, 8-3
Scientific model building, role of statistician, 10-1
Scott and Smith's *t* approximation, 4-40, 4-43
Sensitivity analysis, 9-3
Separation of product variability and imprecision, 2-11
Several or many outliers, 3-26
Shukla's three-instrument imprecision bounds, 2-39
Significance levels for multiple tests, 4-57
Significance tests for three instruments, 2-29
    Bradley-Brindley, 2-50
Significance tests for two instruments, 2-17, 2-23
Single-instrument measurements, 2-8
Skewness test, 3-37
Smallest observation, distribution of, 7-5
Smallest observation, test of, 3-18
Snedecor *F* test, 4-26
Standard deviation, 4-6
Statistical model building, 10-7
Statistical vs physical modeling, 10-1
    Comparison of, 10-11
Statistician's role, 10-1
Statistics of extremes, 7-21
Student's *t*, 4-43
    distribution, 4-40
    percentiles, 4-42
    test (normal approximation—Scott and Smith), 4-41
t test, Student's, 4-40, 6-9
Test for equality of multivariate normal population parameters, 11-4
Tests for highest or lowest observations, 3-12
Thomas-Gart tables, 5-16
Three-instrument case, 2-20, 6A-1
Three-instrument significance tests, 2-29
Tietjen-Moore outlier tests, 3-26
Tolerance intervals, 7-25
    Tolerance factors for normal distribution, 7-33
Transformation of data, least squares, 6-12
True value, 2-6
Truncated Poisson samples, 7-49
Truncated samples, 7-43
Two closest of three observations, 3-9
Two highest or two lowest observation tests, 3-21
2 × 2 contingency tables, 5-9
Unbiased estimation of standard deviation, 4-7
Up and down test strategy, 9-7
Variance, 4-6
Wilk-Shapiro (test of nonnormality), 3-54
Wilks' tests for equality of means, equality of variances, and equality of covariances, 11-4

The proponent of this pamphlet is the US Army Materiel Development and Readiness Command. Users are invited to send comments and suggested improvements on DA Form 2028 (Recommended Changes to Publications and Blank Forms) to Commander, DARCOM (DRCMT), 5001 Eisenhower Avenue, Alexandria, VA 22333.

FOR THE COMMANDER:

OFFICIAL:

CLAUDE M. KICKLIGHTER
Major General, USA
Chief of Staff

for DONNA H. BARTHLE 1LT, GS Asst HQ Adj
CPT, GS
HQ Adjutant

DISTRIBUTION:
Special

ENGINEERING DESIGN HANDBOOKS

These Handbooks are available to Department of the Army activities by submitting an official requisition form (DA Form 17, 17 Jan 70) directly to the Commander, Letterkenny Army Depot, ATTN: SDSLE–SAAD, Chambersburg, PA 17201. "Need to Know" justification must accompany requests for classified Handbooks. Requestors—DOD, Navy, Air Force, Marine Corps, nonmilitary Government agencies, contractors, private industry, individuals, universities, and others—who are registered with the Defense Technical Information Center (DTIC) and have a National Technical Information Service (NTIS) deposit account may obtain these Handbooks from the DTIC. To obtain classified documents from the DTIC, "Need to Know" must be established by the submission of DD Form 1540, 1 Jul 71. Requestors, not part of the Department of the Army nor registered with the DTIC, may purchase unclassified Handbooks from the National Technical Information Service, Department of Commerce, Springfield, VA 22161. NTIS will also accept phone orders (703-487-4650) from holders of the following accounts: Master Charge, VISA, American Express, or a NTIS deposit account. All Handbooks carry the prefix AMCP 706- unless otherwise indicated.

Handbook No. 706-	Document No. (Used by DTIC and NTIS)	Title	NTIS Price
100	AD-890 839	** Design Guidance for Producibility	$28.50
101†	AD-A055 007	Army Weapon Systems Analysis, Part One	$23.50
102†	AD-A086 388	Army Weapon Systems Analysis, Part Two	$39.00
103†	——	Selected Topics in Experimental Statistics With Army Applications	—
106	AD-830 272	Elements of Armament Engineering, Part One, Sources of Energy	$12.00
107	AD-830 287	Elements of Armament Engineering, Part Two, Ballistics	$19.50
108	AD-830 288	Elements of Armament Engineering, Part Three, Weapon Systems and Components	$18.00
109	AD-903 967	Tables of the Cumulative Binomial Probabilities	$45.00
110	AD-865 421	Experimental Statistics, Section 1, Basic Concepts and Analysis of Measurement Data	$16.50
111	AD-865 422	Experimental Statistics, Section 2, Analysis of Enumerative and Classificatory Data	$ 9.00
112	AD-865 423	Experimental Statistics, Section 3, Planning and Analysis of Comparative Experiments	$10.50
113	AD-865 424	Experimental Statistics, Section 4, Special Topics	$12.00
114	AD-865 425	Experimental Statistics, Section 5, Tables	$10.50
115	AD-784 999	Environmental Series, Part One, Basic Environmental Concepts	$18.00
116	AD-A012 648	Environmental Series, Part Two, Natural Environmental Factors	$28.50
117	AD-A023 512	Environmental Series, Part Three, Induced Environmental Factors	$21.50
118	AD-A015 179	Environmental Series, Part Four, Life Cycle Environments	$21.50
119	AD-A015 180	Environmental Series, Part Five, Glossary of Environmental Terms	$17.50
120	AD-889 588	Criteria for Environmental Control of Mobile Systems wC1	$16.50
121	AD-901 533	Packaging and Pack Engineering wC1	$28.50
123	AD-884 519	Hydraulic Fluids	$22.50
124	AD-A025 665	Reliable Military Electronics	$19.50
125	AD-865 109	** Wiring Techniques for Ground and Airborne Systems	$22.50
127	AD-763 495	Infrared Military Systems, Part One	$39.00
128 (S)	AD-C001 857	Infrared Military Systems, Part Two (U)	—
130	AD-830 262	Design for Air Transport and Airdrop of Materiel	$25.50
132	AD-A021 390	Maintenance Engineering Techniques (MET)	$25.50
133	AD-A026 006	Maintainability Engineering Theory and Practice	$25.50
134	AD-754 202	** Maintainability Guide for Design	$31.50
140	AD-830 264	Trajectories, Differential Effects, and Data for Projectiles	$ 7.00
150	AD-462 060	Interior Ballistics of Guns	$15.00
158†	——	* Dynamics of Ballistic Impact, Part One	—
159 (S)†	——	* Dynamics of Ballistic Impact, Part Two (U)	—
160	AD-389 219	Elements of Terminal Ballistics, Part One, Kill Mechanisms and Vulnerability	$15.00
161	AD-289 318	Elements of Terminal Ballistics, Part Two, Collection and Analysis of Data Concerning Targets	$15.00
162 (SRD)	AD-389 372	Elements of Terminal Ballistics, Part Three, Application to Missile and Space Targets (U)	—
163 (C)†	AD-C013 060	Basic Target Vulnerability (U)	—
165	AD-853 719	Liquid-Filled Projectile Design	$14.00
170 (S)	AD-530 922	Armor and Its Applications (U)	—
175	AD-830 265	Solid Propellants, Part One	$13.50
177	AD-764 340	** Properties of Explosives of Military Interest	$30.00
179	AD-777 482	Explosive Trains	$19.50
180	AD-900 260	Principles of Explosive Behavior	$28.50
181	AD-A003 817	Explosions in Air, Part One	$10.50
182 (SRD)	AD-C004 269	Explosions in Air, Part Two (U)	—
185	AD-817 071	Military Pyrotechnics, Part One, Theory and Applications	$21.00
186	AD-830 371	Military Pyrotechnics, Part Two, Procedures and Glossary	$ 9.00
187	AD-830 394	Military Pyrotechnics, Part Three, Properties of Materials Used in Pyrotechnic Compositions	$27.00
188	AD-A000 821	Military Pyrotechnics, Part Four, Design of Ammunition for Pyrotechnic Effects	$22.50
189	AD-803 864	Military Pyrotechnics, Part Five, Bibliography	$10.50
191	AD-884 151	System Analysis and Cost-Effectiveness	$30.00
192	AD-767 826	Computer-Aided Design of Mechanical Systems	$25.50
193†	AD-A055 008	Computer-Aided Design of Mechanical Systems, Part Two	$23.50
196	AD-A027 370	Development Guide for Reliability, Part Two, Design for Reliability	$12.50
197	AD-A032 105	Development Guide for Reliability, Part Three, Reliability Prediction wC1	$15.50
198	AD-A027 371	Development Guide for Reliability, Part Four, Reliability Measurement	$15.00
199	——	* Development Guide for Reliability, Part Five, Contracting for Reliability	—
200	AD-A027 372	Development Guide for Reliability, Part Six, Mathematical Appendix and Glossary	$12.50
201	AD-A022 007	Helicopter Engineering, Part One, Preliminary Design wC1	$21.50
202	AD-A033 216	Helicopter Engineering, Part Two, Detail Design	$23.50
203	AD-901 657	** Helicopter Engineering, Part Three, Qualification Assurance	$25.50
204	AD-785 000	Helicopter Performance Testing	$21.00
205	AD-A020 020	Timing Systems and Components	$21.50

*UNDER PREPARATION—not available
**REVISION UNDER PREPARATION

† DARCOM-P 706-

[over]

☆ U.S. GOVERNMENT PRINTING OFFICE : 1984 O - 442-259

ENGINEERING DESIGN HANDBOOKS (cont'd)

Handbook No. 706-	Document No. (Used by DTIC and NTIS)	Title	NTIS Price
210	AD-889 245	** Fuzes	$18.00
211	AD-389 295	Fuzes, Proximity, Electrical, Part One	$ 4.75
212 (S)	AD-389 331	Fuzes, Proximity, Electrical, Part Two (U)	–
213 (S)	AD-389 330	Fuzes, Proximity, Electrical, Part Three (U)	–
214 (S)	AD-389 333	Fuzes, Proximity, Electrical, Part Four (U)	–
215	AD-389 296	Fuzes, Proximity, Electrical, Part Five	$ 6.25
235	AD-894 910	Hardening Weapon Systems Against RF Energy	$21.00
238	AD-A023 513	Recoilless Rifle Weapon Systems	$23.50
240 (C)	AD-386 896	Grenades (U)	–
242	AD-801 509	Design for Control of Projectile Flight Characteristics	$13.50
244	AD-830 290	Ammunition, Section 1, Artillery Ammunition—General, with Table of Contents, Glossary, and Index for Series	$10.50
245	AD-389 304	Ammunition, Section 2, Design for Terminal Effects	$12.50
247	AD-830 296	Ammunition, Section 4, Design for Projection	$16.50
248	AD-830 284	Ammunition, Section 5, Inspection Aspects of Artillery Ammuntion Design	$ 7.50
249	AD-830 266	Ammunition, Section 6, Manufacture of Metallic Components of Artillery Ammunition	$ 9.00
250	AD-830 303	Guns—General	$10.50
251	AD-838 748	Muzzle Devices	$13.50
252†	AD-830 297	** Gun Tubes	$12.00
253†	AD-A079 666	Breech Mechanism Design	–
255	AD-818 532	Spectral Characteristics of Muzzle Flash	$12.00
260	AD-868 578	Automatic Weapons	$25.50
270	AD-A016 716	Propellant Actuated Devices	$17.50
280	AD-840 582	** Design of Aerodynamically Stabilized Free Rockets	$18.00
281 (SRD)	AD-389 352	Weapon System Effectiveness (U)	–
283	AD-830 377	Aerodynamics	$10.50
284	AD-389 298	Trajectories	$12.00
285	AD-861 082	Elements of Aircraft and Missile Propulsion	$51.00
286	AD-830 267	Structures	$10.50
290†	AD-501 329	Warheads—General	$12.50
298†	——	* Rocket and Missile Container Engineering Guide	–
312	AD-A013 178	Rotational Molding of Plastic Powders	$19.50
313†	AD-A015 181	Short Fiber Plastic Base Composites	$17.50
314†	——	Discontinuous Fiberglass Reinforced Thermoplastics	–
315†	AD-A074 139	Dielectric Embedding of Electrical or Electronic Components	$12.50
316†	AD-A072 362	Joining of Advanced Composites	$17.50
317†	——	* Continuous Fiber Reinforced Composites	–
318†	——	* Plastic Material Properties for Engineering Design	–
319†	——	* Design Considerations of NDE Structural Composites	–
327	AD-830 809	Fire Control Systems—General	$36.00
329	AD-879 465	Fire Control Computing Systems	$34.50
331	AD-830 275	Compensating Elements	$12.00
340	AD-830 276	Carriages and Mounts—General	$ 7.50
341	AD-830 277	Cradles	$ 9.00
342	AD-830 281	** Recoil Systems	$10.50
343	AD-830 393	Top Carriages	$ 7.50
344	AD-830 396	Bottom Carriages	$ 7.50
345	AD-A003 347	Equilibrators	$ 7.50
346	AD-830 301	Elevating Mechanisms	$ 7.50
347†	AD-830 291	Traversing Mechanisms	$ 7.50
348†	——	* Design of Towed Artillery Weapon Systems	–
350	AD-881 357	Wheeled Amphibians	$30.00
355	AD-830 268	The Automotive Assembly	$36.00
356	AD-817 023	Automotive Suspensions	$33.00
357†	AD-873 103	Automotive Bodies and Hulls	$30.00
358†	AD-A035 143	Analysis and Design of Automotive Brake Systems	$15.00
360	AD-783 697	Military Vehicle Electrical Systems	$40.50
361†	AD-A013 769	Military Vehicle Power Plant Cooling	$20.00
410†	AD-A038 803	Electromagnetic Compatibility (EMC)	$22.50
411 (S)	AD-C008 827	Vulnerability of Communication-Electronic and Electrooptical Systems (Except Guided Missiles) to Electronic Warfare, Part One, Introduction and General Approach to Electronic Warfare Vulnerability (U)	–
412 (C)	AD-C008 828	Part Two, Electronic Warfare Vulnerability of Tactical Communications (U)	–
413 (S)	AD-C008 829	Part Three, Electronic Warfare Vulnerability of Ground-Based and Airborne Surveillance and Target Acquisition Radars (U)	–
414 (S)	AD-C008 830	Part Four, Electronic Warfare Vulnerability of Avionics (U)	–
415 (S)	AD-C008 831	Part Five, Optical/Electronic Warfare Vulnerability of Electrooptic Systems (U)	–
416 (S)	AD-C008 832	Part Six, Electronic Warfare Vulnerability of Satellite Communications (U)	–
417 (S)†	AD-C018 700	Vulnerability of Guided Missile Systems to Electronic Warfare (U)	–
430†	——	* Material Deterioration Prevention and Control Guide for Army Materiel, Part One, Metals	–
431†	——	* Material Deterioration Prevention and Control Guide for Army Materiel, Part Two, Nonmetals	–
445†	AD-903 789	Sabot Technology Engineering	$22.50
470†	AD-A029 902	Metric Conversion Guide	$13.50
475†	——	* Quality Engineering	–
480†	——	* Safety Engineering Design Guide for Army Materiel	–

*UNDER PREPARATION—not available
**REVISION UNDER PREPARATION

† DARCOM-P 706-

U238870