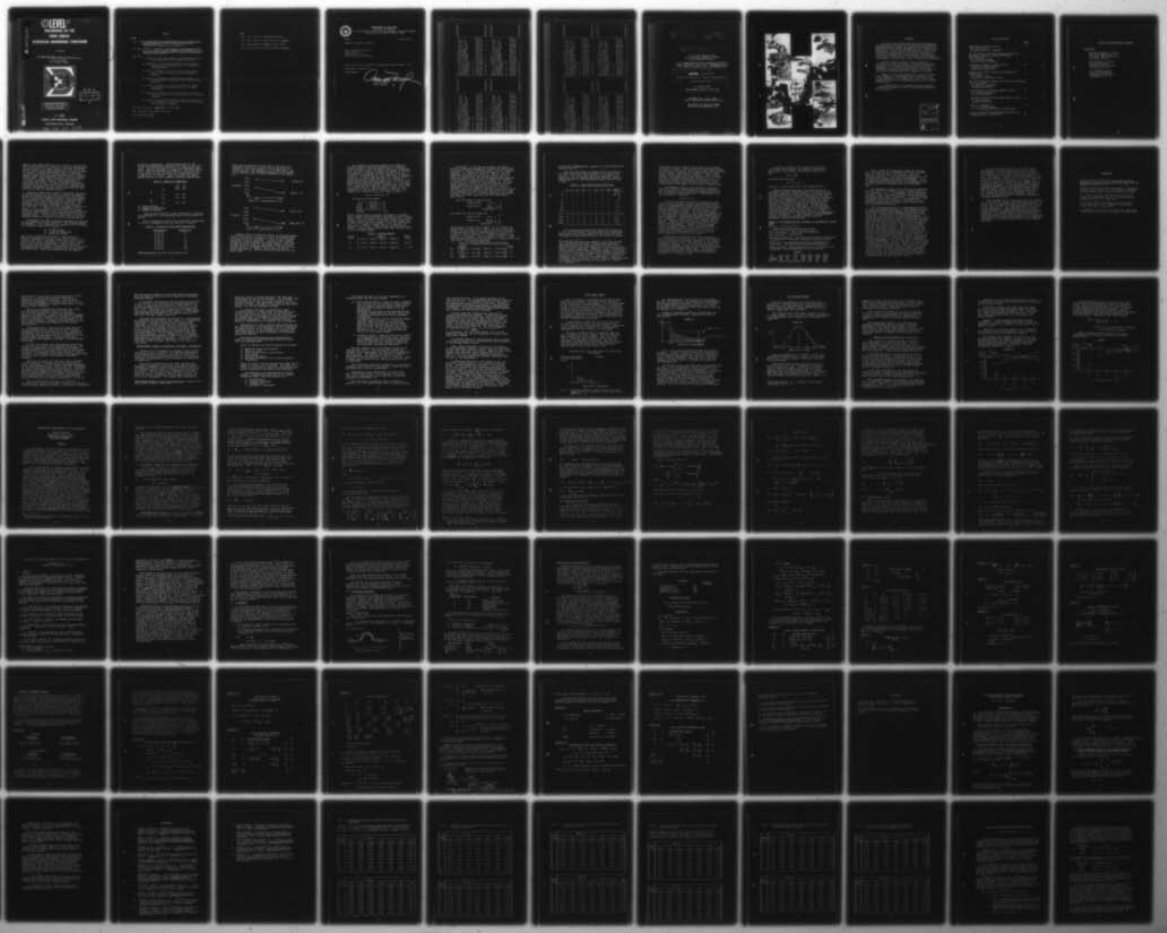
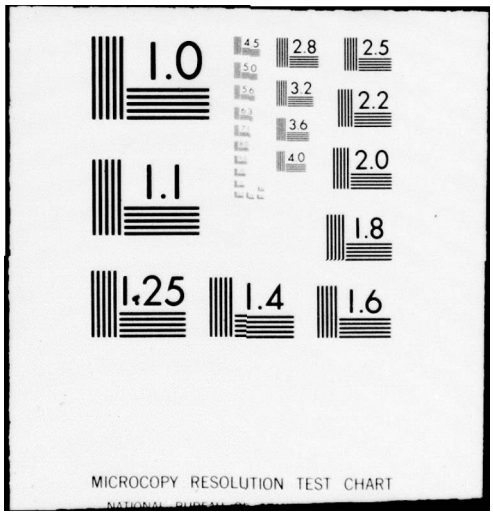
AD-A074 150 ARMY CHEMICAL CORPS ENGINEERING COMMAND ARMY CHEMICAL--ETC F/G 12/1
PROCEEDINGS OF THE ANNUAL STATISTICAL ENGINEERING SYMPOSIUM (3R--ETC(U)
MAY 57 D R HOWES

UNCLASSIFIED NL

1 OF 2
ADA 074150




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART

AD A074150

① LEVEL III

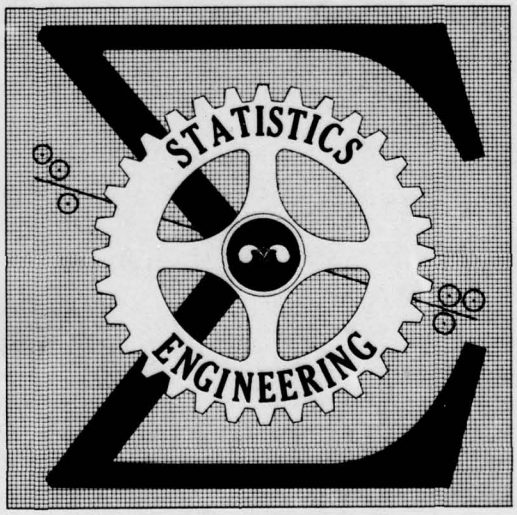
Chem

PROCEEDINGS OF THE THIRD ANNUAL STATISTICAL ENGINEERING SYMPOSIUM

AD-106844

2 - 3 May 1957

US ARMY INFANTRY HUMAN RESEARCH UNIT
P. O. Box 2086
Fort Benning, Georgia



DDC
RECEIVED
SEP 24 1979
B

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DDC FILE COPY

U. S. ARMY
CHEMICAL CORPS ENGINEERING COMMAND
Army Chemical Center, Maryland

79 09 24 356

Errata

Page

ii Add: Mr. Henry Ellner, U. S. Army Chemical Center and Chemical Corps Materiel Command below "Linear Structural Relationships Underlying the Decomposition of Levinstein H"

29 Par. 2, Line 3.- Substitute Biblioabstract on Instrumentation Error by J. E. Doolittle, Report No. R56GL233 BA-135 September 1, 1956 General Electric Laboratory in lieu of "enclosed biblioabstract".

103 Add:

7. Schultz, H., "The Standard Error of a Forecast from a Curve", J. Am. Stat. Assn., Vol. 25 (1930), pp. 139-185

8. Shewhart, W. A., Statistical Method from the Viewpoint of Quality Control. Graduate School, U. S. Department of Agriculture (1939).

9. Wald, A., "Setting of Tolerance Limits When the Sample is Large", Annals of Math. Stat., Vol. 13 (1942), pp. 389-399.

10. Wald, A., "An Extension of Wilks' Method of Setting Tolerance Limits", Annals Math. Stat., Vol. 14 (1943), pp. 45-55.

11. Wald, A. & Wolfowitz, J., "Tolerance Limits for a Normal Distribution", Annals Math. Stat., Vol. 17 (1946), pp. 208-215.

12. Wilks, S. S., "Determination of Sample Sizes for Setting Tolerance Limits", Annals of Math. Stat., Vol. 12 (1941), pp. 91-96.

13. Wilks, S. S., "Statistical Prediction with Special Reference to the Problems of Tolerance Limits", Annals Math. Stat., Vol. 13 (1942), pp. 400-409.

107 Par. 2, Line 2.- Change page 51 to page 116.

109 Add % sign to Iron, Purity and Acid.

114 Change Ns_y to Ns^2_y

Page

115 Par. 1, Line 7.- Insert 116 after page.

115 Par. 4, Line 7.- Change degradate to degrade.

117 Change Raolt's to Raoult's (appears twice).

118 Par. 1, Line 10.- Change obelated to chelated.

DEPARTMENT OF THE ARMY
ARI FIELD UNIT, BENNING
U. S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
P.O. BOX 2086, FORT BENNING, GEORGIA 31905

PERI-IJ

8 August 1979

SUBJECT: Shipment of Documents

Defense Documentation Center
Cameron Station
Alexandria, VA 22314
ATTN: Selection & Cataloging

The Documents in these shipments are approved for public release. The distribution is unlimited.

FOR THE CHIEF:

ALEXANDER NICOLINI
Major, Infantry
R&D Coordinator

AD_____ Accession No._____ | UNCLASSIFIED

Deputy for Engineering, Statistical Engineering Unit, U. S. Army Cml C Eng Cmd, Army Chemical Center, Maryland
PROCEEDINGS OF THE THIRD ANNUAL STATISTICAL ENGINEERING SYMPOSIUM - Edited by Mr. T. M. Vining

May 1957, 124 pp

Proceedings include nine papers which were presented at the Third Annual Statistical Engineering Symposium, held at Army Chemical Center, Maryland on 2 and 3 May 1957.

Abstract Card No._____ | UNCLASSIFIED

AD_____ Accession No._____ | UNCLASSIFIED

Deputy for Engineering, Statistical Engineering Unit, U. S. Army Cml C Eng Cmd, Army Chemical Center, Maryland
PROCEEDINGS OF THE THIRD ANNUAL STATISTICAL ENGINEERING SYMPOSIUM - Edited by Mr. T. M. Vining

May 1957, 124 pp

Proceedings include nine papers which were presented at the Third Annual Statistical Engineering Symposium, held at Army Chemical Center, Maryland on 2 and 3 May 1957.

Abstract Card No._____ | UNCLASSIFIED

AD_____ Accession No._____ | UNCLASSIFIED

Deputy for Engineering, Statistical Engineering Unit, U. S. Army Cml C Eng Cmd, Army Chemical Center, Maryland
PROCEEDINGS OF THE THIRD ANNUAL STATISTICAL ENGINEERING SYMPOSIUM - Edited by Mr. T. M. Vining

May 1957, 124 pp

Proceedings include nine papers which were presented at the Third Annual Statistical Engineering Symposium, held at Army Chemical Center, Maryland on 2 and 3 May 1957.

Abstract Card No._____ | UNCLASSIFIED

AD_____ Accession No._____ | UNCLASSIFIED

Deputy for Engineering, Statistical Engineering Unit, U. S. Army Cml C Eng Cmd, Army Chemical Center, Maryland
PROCEEDINGS OF THE THIRD ANNUAL STATISTICAL ENGINEERING SYMPOSIUM - Edited by Mr. T. M. Vining

May 1957, 124 pp

Proceedings include nine papers which were presented at the Third Annual Statistical Engineering Symposium, held at Army Chemical Center, Maryland on 2 and 3 May 1957.

Abstract Card No._____ | UNCLASSIFIED

AD_____ Accession No._____
Deputy for Engineering, Statistical Engineering Unit, U. S. Army Cml C Eng Cmd, Army Chemical Center, Maryland
PROCEEDINGS OF THE THIRD ANNUAL STATISTICAL ENGINEERING SYMPOSIUM - Edited by Mr. T. M. Vining

May 1957, 124 pp

Proceedings include nine papers which were presented at the Third Annual Statistical Engineering Symposium, held at Army Chemical Center, Maryland on 2 and 3 May 1957.

Abstract Card No._____

UNCLASSIFIED

UNCLASSIFIED

AD_____ Accession No._____
Deputy for Engineering, Statistical Engineering Unit, U. S. Army Cml C Eng Cmd, Army Chemical Center, Maryland
PROCEEDINGS OF THE THIRD ANNUAL STATISTICAL ENGINEERING SYMPOSIUM - Edited by Mr. T. M. Vining

May 1957, 124 pp

Proceedings include nine papers which were presented at the Third Annual Statistical Engineering Symposium, held at Army Chemical Center, Maryland on 2 and 3 May 1957.

Abstract Card No._____

UNCLASSIFIED

UNCLASSIFIED

AD_____ Accession No._____
Deputy for Engineering, Statistical Engineering Unit, U. S. Army Cml C Eng Cmd, Army Chemical Center, Maryland
PROCEEDINGS OF THE THIRD ANNUAL STATISTICAL ENGINEERING SYMPOSIUM - Edited by Mr. T. M. Vining

May 1957, 124 pp

Proceedings include nine papers which were presented at the Third Annual Statistical Engineering Symposium, held at Army Chemical Center, Maryland on 2 and 3 May 1957.

Abstract Card No._____

UNCLASSIFIED

UNCLASSIFIED

AD_____ Accession No._____
Deputy for Engineering, Statistical Engineering Unit, U. S. Army Cml C Eng Cmd, Army Chemical Center, Maryland
PROCEEDINGS OF THE THIRD ANNUAL STATISTICAL ENGINEERING SYMPOSIUM - Edited by Mr. T. M. Vining

May 1957, 124 pp

Proceedings include nine papers which were presented at the Third Annual Statistical Engineering Symposium, held at Army Chemical Center, Maryland on 2 and 3 May 1957.

Abstract Card No._____

UNCLASSIFIED

UNCLASSIFIED

(11) 3 May 57

(12) 139 p.

(10) David R. Howes

U. S. Army Chemical Corps
Major General William M. Creasy
Chief Chemical Officer

U. S. Army Chemical Corps Engineering Command
Colonel William J. Allen, Jr., Commanding Officer
Mr. Louis E. Garono, Deputy for Engineering

(6) Proceedings of the

[redacted] ANNUAL

STATISTICAL ENGINEERING SYMPOSIUM (3rd)

2 - 3 May 1957

Army Chemical Center, Maryland

Chairman: Mr. T. M. Vining
Chief, Statistical Engineering Unit

Secretary: Mr. David R. Howes
Statistical Engineering Unit

036800



INFORMAL ASSOCIATIONS BREED UNDERSTANDING

FOREWORD

The Third Annual Statistical Engineering Symposium was sponsored on 2-3 May 1957 by the U. S. Army Chemical Corps Engineering Command to continue the advancement of knowledge in the rapidly expanding field of "Statistical Engineering". Over a period of three years, thirty-five significant applications of statistical techniques to engineering problems have been presented and discussed by engineers, statisticians and technical administrators representing virtually every type of technical establishment.

The nine papers included in this volume are illustrative of the many types of engineering problems that have been resolved through the use of statistical techniques. It is equally interesting to note the different statistical techniques used by the authors of each paper in approaching their respective problems.

Readers of these proceedings are invited to submit their comments on the program and the administration of the symposium. The sponsor also would welcome the submission of papers for presentation at future symposia of this type.

No reproduction of the papers contained herein is authorized in whole or in part, without the permission of the authors.

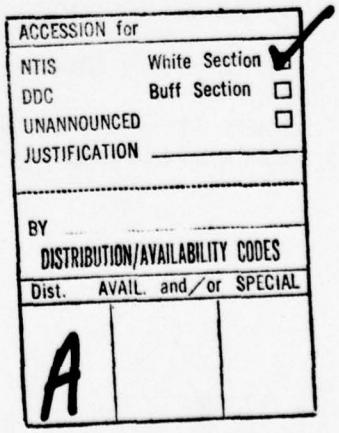


TABLE OF CONTENTS

	Page
Experiments with Many Factors Mr. Marvin Zelen National Bureau of Standards	1
The Derivation of Standard Industrial Ratios of Instrument Accuracy to Design Tolerance Specifications Mr. Leonard Janofsky General Electric Company	13
Statistical Methods in Life Testing Professor Benjamin Epstein Wayne State University	30
Analysis of Variance Models with Engineering Applications Mrs. Mary D. Lum Wright Air Development Center	44
Two-Sided Tolerance Limits for Normal Distributions Using the Range Dr. George J. Resnikoff Stanford University	67
On the Choice of Sampling Inspection Plans Dr. Donald Guthrie, Jr. Stanford University	81
Tightened Multi-Level Continuous Sampling Plans Dr. Cyrus Derman Columbia University	85
Time as a Response Mr. G. Stanley Woodson U. S. Army Chemical Warfare Laboratories	98
Linear Structural Relationships Underlying the Decomposition of Levinstein H	107

STATISTICAL ENGINEERING SYMPOSIUM

Moderators:

Professor Acheson J. Duncan
School of Engineering
The Johns Hopkins University
Baltimore, Maryland

Mrs. Dorothy M. Gilford
Statistics Branch
Office of Naval Research
Washington 25, D. C.

Dr. Joseph Greenwood
Bureau of Aeronautics
Department of the Navy
Washington 25, D. C.

STATISTICAL ENGINEERING SYMPOSIUM

Speakers:

Mr. George B. Beitzel*
Assistant Director for Production
Office of Defense Mobilization
Executive Office Building
Washington, D. C.

Professor Cyrus Derman
Department of Industrial and Management Engineering
Columbia University
New York 27, New York

Mr. Henry Ellner
U. S. Army Chemical Corps Quality Assurance Technical Agency
Army Chemical Center, Maryland

Professor Benjamin Epstein
Wayne State University
Detroit, Michigan

Mr. Donald Guthrie, Jr.
Applied Mathematics and Statistics Laboratory
Stanford University
Stanford, California

Mr. Leonard Janofsky
Missile and Ordnance Systems Department
General Electric Company
Philadelphia 4, Pennsylvania

Mrs. Mary D. Lum
Directorate of Research
Wright Air Development Center
Wright-Patterson Air Force Base, Ohio

Mr. L. Kent Reitz*
Paper Service Division
Eastman Kodak Company
Rochester 4, New York

* Denotes speakers whose papers are not included in these proceedings.

Speakers: (continued)

Professor George L. Resnikoff
Illinois Institute of Technology
Chicago, Illinois

Mr. Leo Tick*
College of Engineering
New York University
40 West 205 Street
New York, New York

Mr. G. Stanley Woodson
U. S. Army Chemical Warfare Laboratories
Army Chemical Center, Maryland

Mr. Marvin Zelen
Statistical Engineering Laboratories
National Bureau of Standards
Washington 25, D. C.

* Denotes speakers whose papers are not included in these proceedings.

EXPERIMENTS WITH MANY FACTORS

By Marvin Zelen
National Bureau of Standards

Among the most difficult types of experiments to conduct is the experiment to investigate the joint effects of several factors on the performance of a piece of equipment, or the yield of a product, or the characteristics of a test method. A considerable advantage is gained if the experiment is conducted so that the effects of changing each variable can be evaluated jointly with the effects of changing the other variables which might conceivably affect the outcome of the experiment. One way of achieving this objective is to decide on a set of conditions for each factor and to carry out one or more measurements for every possible combination of the conditions. Such experiments are termed factorial experiments or multi-factor experiments.

Other things being equal, the smaller the number of factors, the fewer the difficulties encountered. However, because of possible complex inter-dependencies among the various factors, the difficulties associated with factorial experiments become formidable for even a moderate number of factors.

The first application of the formal theory of factorial experimentation was devised to fit the special problems of agricultural research. Here, the agricultural scientist planted in the spring, harvested in the fall and used the entire winter for analyzing the results. Somewhat later, efforts were made to adapt these experimental procedures for use in the physical and engineering sciences. The recent growth of the subject of the statistical design of experiments has occurred in response to the need for methods specially suited to these new areas of applications.

Among the difficulties encountered with the application of factorial experiments in the physical and engineering sciences is that the total number of different combinations of conditions may be quite large and in many cases prohibitive. Another disadvantage is that in many experimental situations it is not practical to plan an entire experimental program in advance, but instead, to make a few

smaller experiments which serve as a guide to further work. This latter condition is especially true when measurements are made singly or in small groups, such that the experimental results become known sequentially as they are taken. In response to these disadvantages ways of conducting factorial experiments have been developed which require a smaller number of measurements selected from all possible combinations. These reduced factorial experiment plans are called fractional factorial experiment plans or fractional replicates. The key idea behind fractional replication is to choose a sub-set from all possible combinations such that the sub-set chosen (i) contains more relevant information than any other sub-set and (ii) the analysis is easy and straight-forward.

Recently a publication entitled "Fractional Factorial Designs for Factors at Two Levels", authored by members of the Statistical Engineering Laboratory of the National Bureau of Standards, appeared in print. This publication is a catalogue of fractional factorial experiment plans covering a wide range of experimental situations. The fractional plans in this catalogue are for situations covering from 5 through 16 factors and enable fractions of 1/2, 1/4, 1/8, ..., 1/256 of a full experiment to be chosen. It is the purpose of my talk today to discuss fractional experimentation and to describe these fractional factorial plans. However, before describing fractional experimentation, I would like to discuss some of the basic concepts involved in complete factorial experimentation.

An example of a simple factorial experiment is provided by an experiment recently conducted at the National Bureau of Standards. This experiment was concerned with evaluating the strength of steel with respect to

A: Carbon Content
B: Tempering Temperature
C: Method of Cooling.

Part of this experiment consisted of considering two different conditions for each of the three factors. It will be convenient to designate these three factors by the letters A, B, and C, respectively. Also we shall denote one of the particular conditions at which a factor is evaluated by the sub-script 0 and the other by a sub-script 1. For example the two tempering temperatures involved were $400^{\circ}F$ and $600^{\circ}F$. These are represented by the symbols

B_0 and B_1, respectively. Statisticians refer to the different experimental conditions of a factor by the term level. Thus each of the factors in this experiment has 2 levels, and the experiment is termed a 2^3 factorial experiment. Often we term the level with the 0 sub-script as a "low-level" and the level with the 1 sub-script as the "high-level" of a factor. Table 1 contains the experimental results for the 8 possible factorial combinations.

Table 1. Results of Experiment

		A_0	A_1
B_0	C_0	169	167
	C_1	173	165
B_1	C_0	145	135
	C_1	143	134

A = Carbon Content
B = Tempering Temperature
C = Method of Cooling

These results [1] (coded for easy presentation) represent the strengths of 8 steel specimens after a given period of stress.

Now in designating each of the 8 factorial combinations, it will be convenient to use the notation of Table 2.

Table 2. Notation for Factorial Combinations

Combination	Designation
A_0 B_0 C_0	(1)
A_1 B_0 C_0	a
A_0 B_1 C_0	b
A_1 B_1 C_0	ab
A_1 B_0 C_1	ac
A_0 B_1 C_1	bc
A_1 B_1 C_1	abc

[1] The measurements are lbs./in^2 divided by 10^3

Note that the notation is such that if the level of a particular factor is at the low level, the letter is missing; if the level present is at the high level, the letter appears. The combination where all three factors are at the low level is simple designated by the number (1).

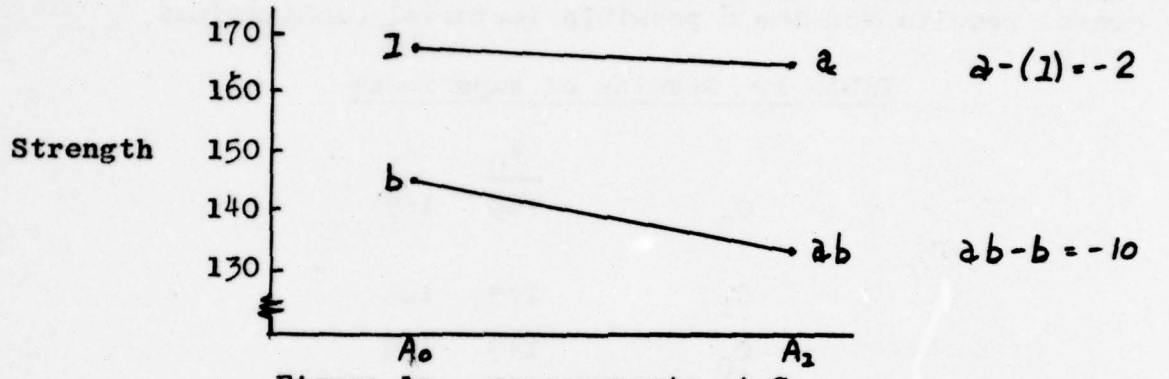


Figure 1a - measurements at C_0

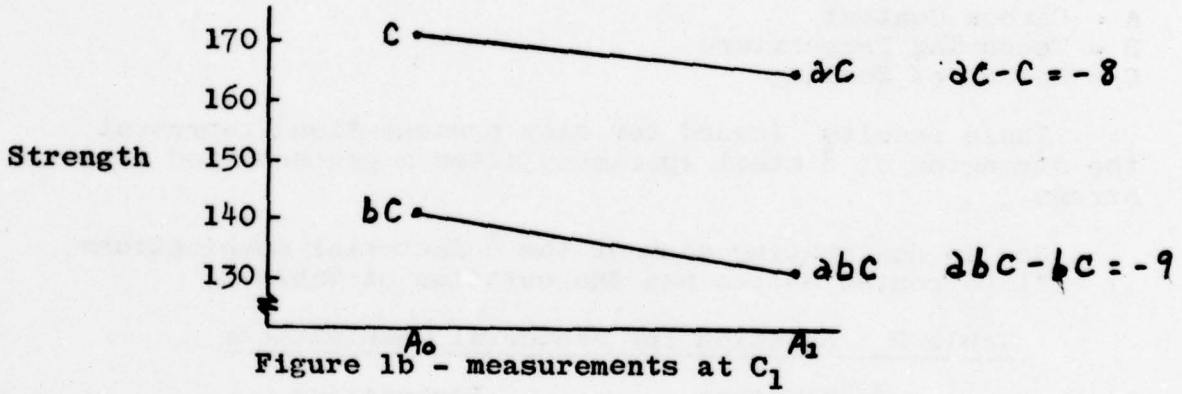


Figure 1b - measurements at C_1

Figures 1a and 1b graph the results of the experimental results where figure 1a refers to the 4 measurements made at the C_0 condition and figure 1b refers to the 4 measurements made at the C_1 condition. In each of these graphs, strength is plotted against the carbon content. The lines joining the pairs of points join those points which have the same level of tempering temperature. That is the line joining (1) and a in figure 1a joins those two points having condition B_0. Similar interpretations hold for the other lines.

Now suppose we wished to evaluate the change of strength resulting from considering the two different carbon contents. Since the carbon factor is represented by A, we could regard the effects of a change in strength with respect to a change in carbon content to be reflected by the slopes of the four lines shown in figures 1a and 1b. If the slopes are near zero, then we would be justified in stating that the effect of a change in carbon content on strength is small over the range of the experimental conditions encountered here. It is an easy matter to calculate the slopes for each of the four lines. If we regard the difference in carbon content to be equal to a unit difference, then the slope is equal to the difference of the two points joined by each line. For example, for condition B_0 C_0, the slope is equal to

$$a-(1) = 167-169 = -2.$$

We have four such slopes

a-(1)	=	167-169	= -2
ab-b	=	135-145	= -10
ac-c	=	165-173	= -8
abc-bc	=	134-143	= -9
Average			= -7.25

Each of these slopes represents the change in strength with a change in carbon content, holding the other factors fixed. Now if one wanted the average change due to carbon content, over the range of conditions of this experiment, we would simply take the average of the four slopes which is = -7.25. Similar calculations can be made for the remaining two factors. These results are summarized in Table 3. Statisticians usually label these average slopes by the term main effects

Table 3. Average Slopes

Factor	Formula	Value
A	$\frac{1}{4}$ (a-1) + (ab-b) + (ac-c) + (abc-bc)	-7.25
B	$\frac{1}{4}$ (b-1) + (ab-a) + (bc-c) + (abc-ac)	-29.25
C	$\frac{1}{4}$ (c-1) + (ac-a) + (bc-b) + (abc-ab)	- .025

From Table 3 it is easy to see that, in general, factor B (tempering temperature) produced the most change in strength. Factor C produced almost a trivial change in strength. Now suppose we wished to evaluate the interdependence of factor A on factor B. That is, does the effect of carbon content on strength depend upon the tempering temperature? One way of answering this question is for fixed C, to compare the two slopes with respect to factor A for condition B_0 and condition B_1. In other words, looking at figure 1a, which represents the four measurements made at condition C_0, we would like to ascertain if the two slopes are parallel. If they are parallel, this would indicate that the change due to factor A is independent of factor B and factor A is said to not interact with factor B. A similar comparison can be made for figure 1b which represents the four measurements made with C_1.

The two slopes on figure 1a are

slope for B_0:	a-(1)	= -2
slope for B_1:	ab-b	= -10
	difference	= 8

For figure 1b, the two slopes are

slope at B_0	ac-c	= -8
slope at B_1	abc-bc	= -9
	difference	= +1

If the pairs of slopes are parallel, we would expect the differences to be near zero. Again, as with the average slopes, we could take the average of these two differences to represent the average interdependence of factor A on factor B. This results is 4.5. Similar calculations can be made for AC and BC. These are shown in Table 4. When these average differences are divided by 2, statisticians term the result a two-factor interaction.

Table 4. Average Differences Between Slopes

	Formula	Value
AB	$\frac{1}{2}\{[(a-1) - (ab-b)] + [(ac-c) - (abc-bc)]\}$	4.5
AC	$\frac{1}{2}\{[(a-1) - (ac-c)] + [(ab-b) - (abc-bc)]\}$	2.5
BC	$\frac{1}{2}\{[(b-1) - (bc-c)] + [(ab-a) - (abc-ac)]\}$	2.5

The greatest interdependency appears to be that between the factors A and B.[1]

Note that the formula in tables 3 and 4 are made up of linear functions of the observations, divided in one case by 4, and in the other case by 2. For purposes of comparison, we can dispense with these divisors. Table 5 brings together all the appropriate formula without these divisors.

Table 5. Main effects and interactions

	1	a	b	ab	c	ac	bc	abc	Value
A	-	+	-	+	-	+	-	+	-29.00
B	-	-	+	+	-	-	+	+	-117.00
C	-	-	-	-	+	+	+	+	.10
AB	-	+	+	-	-	+	+	-	9.0
AC	-	+	-	+	+	-	+	-	5.0
BC	-	-	+	+	+	+	-	-	5.0
ABC	+	-	-	+	-	+	+	-	-7.0
A+BC	-2	0	0	+2	0	+2	-2	0	-24.0
B+AC	-2	0	0	+2	0	-2	+2	0	-108.0
C+AB	-2	0	0	-2	0	+2	+2	0	9.1

The quantities associated with the single letters A, B, and C are multiples of the main effect or average slope. These measure the trends produced by the various factors. The quantities associated with pairs of letters are simply

[1] If this same experiment was completely run again and the same calculations were made, we would expect different numerical results than those presented in Tables 3 and 4. These different results are simply due to the random errors inherent in the experiment. Therefore inorder to objectively decide whether certain apparent effects amongst the factors are real one would have to compare the numerical results with a measure of the random errors. There are many statistical techniques for carrying out such comparisons. These are classified under the general title of "Tests of hypothesis" and are somewhat beyond the scope of this paper, cf. Davies (3), Kempthorne (5).

a multiple of the two factor interaction or the average difference in slope. These measure the dependence of pairs of factors on each other. By an easy formal generalization, one can define a three-factor interaction which is presented (without the divisor) in Table 5. Also, if more factors were involved, we could formally define 4-factor, 5-factor, and other higher order interaction terms. In this paper we shall be confining our attention to main effects and 2-factor interactions as the higher order interactions do not appear to arise often in applications.

Now suppose we were in the situation of wanting to select half of the 8 combinations for experimentation. For this purpose, let the four measurements selected for the 1/2 fraction be those having a + coefficient in the ABC interaction. These are

1, ab, ac, bc.

If we had run this 1/2 replicate experiment, aside from a loss in precision arising from taking a smaller number of measurements, another penalty involved would be that the estimates of the main effects and interactions become entangled or "aliased" with one another. For example, referring to Table 5, we can add the formula for the main effect A to BC. This gives the estimate for A + BC which only involves the four measurements used in the 1/2 replicate. This formula appears in the bottom half of Table 5. Usually we say that the main effect A and the two-factor interaction BC are aliased with one another. There is no way of separating the two using these four measurements. However, if from a priori reason, we could consider the BC interaction to be zero or negligible, then we might interpret A + BC as reflecting principally the main effect A. Similar quantities can be calculated for B + AC and C + AB which appear in the lower part of Table 5.

Previously we pointed out that the measurements selected for the 1/2 replicate were those which appeared with + coefficients in the ABC interaction. We usually indicate this by I = ABC. This quantity is called the fundamental identity. It can be used for selecting the fractional replicate and also for determining the aliases of all main effects and interactions. The factorial combinations were simply those combinations having an even number of letters in common with the fundamental identity, i.e., (1), ab, ac, bc.

In order to determine the aliases of any factor, we multiply both sides of the fundamental identity by that factor. For example, for the main effect A we have

$$A \cdot I = A^2BC$$

and using the convention

$$A \cdot I = A, \quad A^2 = 1$$

we have A = BC and state that A is aliased with BC.

These principles represent the basic aspects of fractional factorial experimentation. For any particular fraction, there will be many different fractional designs. The National Bureau of Standards designs have been constructed on the premise that the most important information needed is that associated with main effects; the second most important information is that on two-factor interactions. Information on higher order interactions is considered of negligible importance. The NBS designs were so made up that all main effects were aliased with 3-factor and higher order interactions. If a two-factor interaction was aliased only with 3-factor or higher order interactions, the interaction was called a measurable 2-factor interaction. It was attempted to construct the designs to attain the maximum possible number of measurable 2-factor interactions.

Table 6. 1/8 replication of 8 factors in 8 blocks of 4 units each.

Factors: A,B,C,D,E,F,G,H,

I - ABEGH-ACFG-BCEFH-ABCD-CDEGH-BDFG-ADEFH.

Block confounding: EGH,FG,EFH,BEH,BG,BEFGH,BF.

Row confounding: ABEF,ACE,BCF.

Completely randomized: The following two-factor interactions are measurable: AE,AH,BE,BH,CE,CH,DE,CH,EF,EG,EH,FH,GH.

Blocks only: The following two-factor interactions are measurable: AE,AH,BE,BH,CE,CH,DE,CH,EF,EG,EH,FH,GH.

Blocks and Rows: The following two-factor interactions are measurable: AE,AH,BE,BH,CE,CH,DE,CH,EF,EG,FG,GH.

Blocks							
1	2	3	4	5	6	7	8
(I)	$a\bar{b}cd$	$a\bar{c}fg$	$bd\bar{f}g$	$cd\bar{e}fh$	$a\bar{b}efh$	$a\bar{d}egh$	$b\bar{c}egh$
abcdefg	efg	bde	ace	abgh	cdgh	bcfh	adfh
abcdfgh	fgh	bdh	ach	abeg	cdeg	bcef	adef
eh	abcdeh	acefgh	bdefgh	cdf	adf	abf	bcg

Table 6 shows the information needed for carrying out a 1/8 fraction of an experiment involving 8 factors such that each factor is at two levels. The full factorial requires $2^8 = 256$ different combinations. The 1/8 fractional design needs only 32 measurements. The 32 selected measurements are given in 8 columns. It can be shown that no other selection of 32 measurements will give more relevant information than this experiment plan.

The fundamental identity for this plan is associated with the letter I as before, however, because this is a 1/8 replicate, $2^3 - 1 = 7$ terms appear in the fundamental identity. Thus, every main effect and interaction will be aliased with 7 other interactions. Of the 28 possible two-factor interactions, exactly 13 are measurable. The interpretation of the remaining 15 two-factor interactions is questionable as these contain aliases which have two-factor interactions.

Often the measurements or the experimental material may come in homogeneous groups. These homogeneous groups of measurements are called blocks. For example, quite often only a limited number of measurements can be made in one day. In this case, the measurements made on a single day would represent a block of measurements. Then if measurements made on a single day show better agreement with each other than measurements from different days, we would like to take this into account in our experimental plan so as to avoid or balance out any biases arising from the day to day variation. These fractional plans have been constructed so as to take this situation into account. The eight columns in this 1/8 replicate show how one would assign the measurements to each block so as to balance out any possible biases arising from this source. The particular measurable 2-factor interactions available, if blocking is used, are listed alongside the heading Blocks only. Here we still have 13 measurable 2-factor interactions. On the other hand, even if we take into account differences between blocks, the order within a block might also be important. This would be the case if the measuring equipment would be subject to a drift during the day. In order to balance out this possible source of bias, the fractional designs have been constructed so that the experimental layout corresponds to the proper order of testing needed to balance out any biases arising from this source of error.

Now I would just like to say a few words about applications. In some situations we might run our experiments in parts so that at the end of any part of the experiment we can go back and easily analyze the data. For example, we might run a 2^8 experiment in sets of 32 measurements. That is, we can run a 1/8 fraction, analyze it; if conditions warrant, run another 1/8 fraction to give a 1/4 replicate, etc. Another use for these fractional plans is in trouble-shooting a piece of complex equipment. In this situation it is usually possible to list a great many possible factors which could go wrong. However, the chances are that only a few of these factors are not functioning correctly. Here fractional designs giving information only on main effects should be useful in making a diagnosis of equipment malfunctioning.

This catalogue is for experiments where all factors are at two levels. However, by an easy modification they can be used for experiments where all factors are at 4 levels, or in general, for experiments where the factors are all powers of 2. Also, sometime later this year another catalogue of fractional designs for factors at three levels will be published by the National Bureau of Standards. It is hoped that with the publication of these catalogues of fractional designs, it will make the application of fractional experimentation more immediately accessible to the applied statistician and the working scientist.

REFERENCES

1. National Bureau of Standards, Fractional Factorial Experimental Designs for Factors at two Levels, Applied Mathematics Series 48, 1957.

2. Cuthbert Daniel, Fractional replication in industrial research, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability V (1956)

3. O. L. Davies (editor), The design and analysis of industrial experiments, Chap. 10 (Hafner Publishing Company, New York, New York, 1954).

4. D. J. Finney, Recent developments in the design of field experiments. III. Fractional replication, J. Agr. Sci. 36, 184-191 (1946)

5. O. Kempthorne, The design and analysis of experiments (John Wiley and Sons, Inc., New York, New York, 1952)

THE DERIVATION OF STANDARD INDUSTRIAL RATIOS OF INSTRUMENT ACCURACY TO DESIGN TOLERANCE SPECIFICATIONS

By Leonard Janofsky

STATEMENT OF OBJECTIVES

At the time an instrument is selected to measure a design specification it is the usual procedure in industry to select an instrument rated by the manufacturer at one-tenth the design tolerance required. This will be referred to throughout this report as the ten-to-one ratio.

Regardless of the scaler quantities involved, this ratio has been applied as a rule-of-thumb. It will be demonstrated below that the application has resulted in an excessively tight tolerance being placed upon instrumentation requirements at the high accuracy levels. As a result there has been considerable addition to instrumentation cost as well as delay in the design and/or procurement of test equipment.

The purpose of this report will be to present a survey on how the ten-to-one ratio of design tolerance to instrument tolerance was derived. Statistical Techniques for the derivation of required new standards will be discussed and contributions to instrument error will be analysed - all with a view toward replacing the ratio with a less expensive and one which is more applicable to high accuracy instrumentation.

THE DERIVATION OF THE TEN-TO-ONE RATIO

The selection of a ten-to-one ratio does not have any strong mathematical basis. Where high accuracy is required design tolerances are normally expressed in decimal form. Therefore, it has been convenient to express instrument accuracy tolerances in a readily convertible form as a multiple of ten.

This conclusion is based upon a complete review of the literature available in this field, both here and abroad. A biblioabstract has been constructed which demonstrates that no mathematical basis for this ratio

was ever derived.[1] In addition, industrial organizations employing this ratio were consulted and their personnel bear out this contention both as to the derivation and applicability of this ratio.

The industry has been particularly concerned with justifying selection of the ten-to-one ratio and several projects have been conducted to do this. For example, Fred Law of the General Electric Co. demonstrated that an error of approximately ten percent was present when mechanical parts were inspected by 150 inspectors at various plants throughout the General Electric Company.

In general, the wide acceptance of this ratio was primarily due to an intuitive justification - i.e. the system provided a consistent measure which could be applied "across the board".

Since the ratio was used simply as a standard, there need be no mathematical significance to any specific ratio. It wouldbe preferable, however, to establish some new ratio based upon the old one in a manner designed to readily convert values and standards derived from the ten-to-one ratio. The function had served industry well so long as certain low levels of accuracy were required. There existed no requirements for a new ratio since instrumentation was not available or required for higher accuracy work.

However, since machine tooling has become capable of maintaining tighter tolerance limits on production units, particularly since World War II, there has been no general adherence to this ten-to-one ratio in industrial practice at the tighter tolerance zones. This has been due partially to the expense and lack of immediate availability of high precision instrumentation. To a much greater extent, this has been due to the reduced requirements for such a tight ratio at the tighter tolerance levels due to factors described below.

Although there has been no general adherence to a specific ratio the need for a modified commercial standard has existed for the past several years.

---

[1] Biblioabstract on Instrument Error - Jane E. Doolittle, Report No. R56GL233BA-135, General Engineering Laboratories, General Electric Company, September 1956.

It would be preferable to relate a new standard to the previously derived ratio for obvious reasons. Any new standard must also be flexible so that the same problem does not arise at some future date when metrology and requirements for accurate instrumentation would make previously collected data incomparable. Finally, a mathematical basis for a new standard would facilitate acceptance in industry.

As a result, considerable demand has been generated recently for new standards of instrumentation. It has been demonstrated in industry that the instrument accuracy called for at tight tolerance levels is not readily available in commercial instruments without excessive expense and time delay. It remains for industry to derive and accept a new standard more compatible with the high accuracy instrumentation required.

THE TREND OF HIGH ACCURACY INSTRUMENTATION

The history of high accuracy measurement is one which originated with the industrial revolution and became particularly important during and after World War II. The derivation of standard industrial ratios of manufactured instrument accuracy to design tolerance specifications closely parallels the history of accurate measurement.

Over two thousand years ago, standards of measurement comparable to those employed in engineering today were obtained. This was primarily due to the fact that factors requiring measurement were extremely rough compared to the instrumentation available.

The modern history of accurate instrumentation originated with James Watt's steam engine. John Wilkinson's boring mill was producing a round cylinder so that a "well worn shilling" wouldn't slip between the piston and bore.

By 1900, British shipyards on the Clyde were employing a tool called the "micro-meter", which (it was claimed) could measure to one-thousandth of an inch.

Prior to 1930 the need for a standardized ratio of instrument tolerance to design tolerance was not felt strongly in the manufacturing areas. The technological development of these measurement instruments was at a

stage which did not require a functioning ratio of less than 10 to 1 - design specification tolerance to instrument manufacturer's tolerance. The reasoning behind the establishment of this the ratio at a 10 to 1 level was the inadequacy of machine tools which, in turn, required loose tolerances.

Most of the technological inovations came in the area of automotive design. About 30 years ago, Col. William d'Armody had developed an automatic hydraulic transmission. However, the hydraulic slippage caused by the loose tolerances and methods of production made it impossible to deliver the required power at the end of the driveshaft. Standard instrumentation had not yet attained the perfection demanded of the design specifications.

During World War II fuel-injection parts and ball bearing assemblies were some of the areas requiring standards of one-five millionth of an inch. Production requirements went to two or even one-ten-thousandth of an inch, forcing gage manufacturers to employ 20 or even 10 millionths on some gages. Reference instrument manufacturers were required to go down to two millionths of an inch.

By 1946 the industry was merely paying lip service to the ten-to-one ratio at the tightest tolerances. The accuracy of measurement and technological development rapidly increased to the stage where industry was working to a tenth of a millionth of an inch by means of interferometry. This method is based upon the precise splitting of light waves to achieve the necessary measurement standards.

The trend in industry at this point was to develop a specialization in the high accuracy instrumentation field. For example, the Eli Whitney Metrology Laboratory of the Sheffield Corporation, Dayton, Ohio was established to achieve this high accuracy instrumentation. At present the Large Steam Turbine and Generator Department of the General Electric Company is developing a device which will air test experimental blade and chamber configurations with a precision of $\pm$ 0.1 percent at pressure ratios of 1.05^{-10} with inlet pressures ranging from 2 to 140 psi.

The best attainable accuracies in electrical instruments have been periodically compared to the accuracy of the ordinary commercial products. In England, considerable

work has been performed by the National Physical Laboratory which provides an indication of the quality of these manufactured products.

The moderate accuracies for British Ammeters specified in B.S. 89 1937 or in the Electricity Supply (meters) Act of 1936 in England were not achieved in 33 percent of the instruments submitted to the N.P.L. for test in the years 1945 - 1950. Since most of the defective instruments were repaired and eventually passed their tests, it would appear that approximately one-half of the ammeters submitted to the N.P.L. for test failed initially.[2]

This has been attributed not so much to a deterioration of the quality of instruments, but rather to a technological improvement in methods of measuring and testing these instruments, indicating that possibly instruments were inherently not of the quality that they were once thought to be. The previous ratings of these instruments were overstated and the applicability of the ten-to-one ratio was established with less accurate instruments than was previously suspected. In all probability, a five-to-one ratio may have been employed to assert proof of industrial applicability of the ten-to-one ratio. A ten-to-one ratio requirement placed upon an instrument in 1930 would constitute a more stringent requirement on the identical test system.

MANUFACTURER'S ERROR IN ELECTRICAL INSTRUMENT CONSTRUCTION

Precision of an instrument is defined as the internal consistancy of the instrument. The complete understanding of the appearance of errors and how they are affected by details of construction and method of use of the measuring instrument is prerequisite to any discussion of reliability of measurement.

Electrical instruments are too often regarded more as pieces of apparatus to be standardized against similar instruments of higher accuracy than as measuring devices to be introduced into a circuit to determine the condition of the circuit. However accurate an instrument may be when tested as a unit, if this introduction into a circuit disturbs the circuit conditions, the instrument reading may be of little consequence. The tighter the tolerances to

[2] Performance Limits in Electrical Instruments - Arnold A.H.M. Proc. I.E.E. 98 Part 2, 701 (1951).

which the circuit is being subjected, the lower the significance of the ten-to-one ratio. The instrument will not disturb the circuit as a linear function of the accuracy required. The disturbance caused by the introduction of a test instrument is a function of the circuit design - not of the instrument accuracy.

In accessing the value of instruments, their limitations must to a large extent be regulated by the circuit conditions into which they may be connected. The influence of the introduction of these errors should be included when an instrument is being selected for a particular measurement. The contributing influence should be considered part of whatever ratio is selected.

The ability of a test to detect small variations in the characteristic to be measured is of basic importance. The characteristic of an instrument that it can produce consistent results is not necessarily a desirable one where measurements of small variations in the measured medium are required.

The defects which distinguish a manufacturer's product and special laboratory models may be divided into two classifications:

Class "A" Defects (not easily detected by the user.)

1) Resistors incorrectly adjusted
2) Errors in scale
3) Power-factor errors in watt meters
4) Level errors
5) Self-heating errors
6) Movements sticking at one or more positions
7) Variability

Class "B" Defects (easily detected but liable to be a source of error if not rectified). The Class "B" defect is a function of level of inspection and subsequent handling of the manufactured instrument.

The precision of the manufacturers' average product is appreciably below that of the special laboratory model - with the exception of

1) Potentiometers
2) Standard resistance
3) Standard cell
4) Instrument transformers

The errors included in electrical measurement are broadly divided into two main classes.

(a) The Accidental Errors. These are due to elements over which we have little direct control such as noise or smoke in the laboratory, vibration of the building and physical conditions of the experimenter.
The physical conditions of the experimenter may be further broken down into fatigue, faulty and incompetent manipulation and lack of practice or knowledge.
Although little direct control can be maintained over this source of error, it is still an important function of the errors which have habitually been excluded from the ten-to-one ratio and might well be considered as a future addition to it. It may be well to derive two additional ratios - one for physical factors and one for the human elements.

(b) Systematic Errors. These are the errors inherent in the equipment and the method used. They are also dependent upon some of the conditions under which the experiment is conducted; conditions which are known and the influence of which can be eliminated, minimized or calculated.

The most important of the systematic errors is the constructional error. This error arises because the equipment can be guaranteed only where certain assumptions are held valid. For example, the effects of frequency, self-inductance, and distributed capacity should be submitted by the manufacturer (either by a formula, graph, or phase angle) from which the constructional error can be calculated.

This information should be provided to all statistical analyses of the results of such testing - along with the specifications of the test conditions.

Inductances, mutual inductances, and capacities, have their constructional error indicated in a similar way, and the same applies to tuning-forks, wavemeters, generators, instruments, etc.

With indicating instruments such as voltmeters, ammeters, etc, the constructional error is rarely uniform

over the entire scale. It is usually greatest in the first third of it. The manufacturer may therefore indicate the error as say, ± 0.5 percent in the rest. Naturally, any derived ratio should be a function of the non-uniform constructional error where high accuracy instrumentation is involved. The errors introduced become proportionately more important as the accuracy increases.

In some cases the constructional error may be indicated by significant figures. A significant figure constructional error would be digit which is thought to be nearer to the "true value" than is any other digit. This is most often used in guarantee certificates. For example, an inductance may be expresses as 125500 mh or 1.255×10^5 mh, which means that the "true value" is between 125400 mh and 125600 mh. The limits expressed as a percent are here $\pm \frac{100}{125500} \times 100 = \pm 0.0797 \cong$ ± 0.08 percent. The first four digits, 1255, are the significant figures, and the inductance is given to four significant figures.

In whatever manner the constructional error is given, it is best converted into a percentage for the calculation of the maximum possible error.

Again, when a resistance is given as 12222 ohms or to five significant figures, it means that its value lies between 12221 ohmes and 12223 ohmes. Zeros are significant figures only when other digits precede them in the number; thus 0.0125 has three significant figures and not five. The more significant figures, the greater the accuracy of the measurement.

A substantial reduction in the amount of error in electrical instrumentation is not expected in the near future. Further substantial advances in the accuracy of the best instruments under favorable conditions are not to be expected. In most cases standard commercial instruments have undergone only detail changes in the past few years. Advances which may be expected will be with a view toward extending the range of conditions for maximum accuracy. The technological developments are expected in the relm of a reduction in the margin of performance between the specially adjusted instrument and the ordinary level of commercial instrumentation. Additional technological improvements can be expected in a reduction of the number and magnitude of the corrections to be applied to an instrument reading.

"TRUE VALUE" ERROR

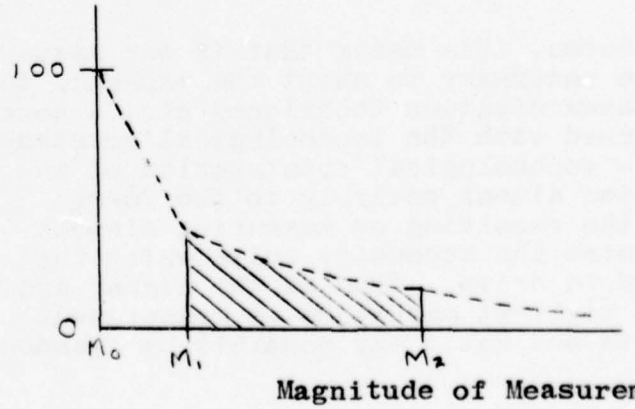
The "true value" of a variable is defined by fixed points or by standards. The "true value", therefore, could conceivably be an approximation. By the replication of the "true value" of the variable and by a statistical review of the consistancy of the results, a function can be derived which is interpreted as a measure of the precision or maximum variation in indication that will occur. The precision function can be expressed as the positive and negative difference between a given measurement and the average indication for a given true value found from several independent determinations.[3]

It can readily be seen that since the "true value" is determined by the quality of the standard which, in turn, depends upon intercomparison methods, uncertainty could conceivably be introduced into the measurement regardless of the basic design of an instrument.

Any error introduced which is dependent upon the basic standards is thus a constant error not dependent upon the magnitude of the reading. The "true value error", therefore, would always be present when any instrument is compared to a standard. This error is negligible where the magnitude of measurement is large, but its contribution to the total error increases significantly at the extreme values, thus:

Contribution of True Value Error to Total Error
(in percent)

FIGURE I

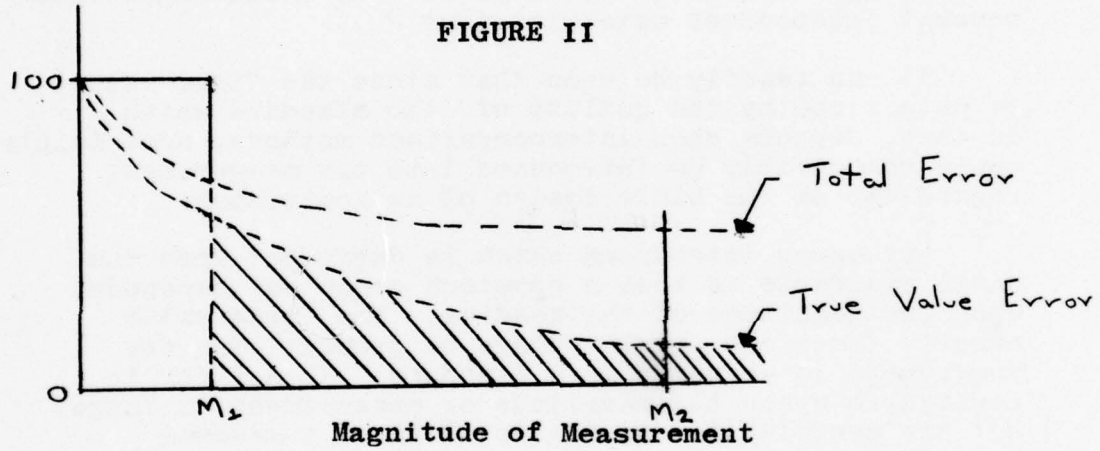


Note 3 - Measurement Errors - Classification and Interpretation J. C. Boonshaft. Transactions of the ASME, May, 1955.

Our discussion will center about the area between M_1 & M_2. The portion M_o - M_1 represents a technologically unattainable degree of accuracy. The area to the right of M_2 represents the area for which the 10 to 1 ratio was derived. The importance of the "true value" error can be demonstrated by Figure II.

Assume an instrument capable of becoming more and more technologically accurate. Measure the error on the ordinate.

FIGURE II



We are still only concerned with the area between M_1 and M_2. However, it will be noted that althought the contribution of this "calibration" error to total error is at a constant percentage, the total error gradually rises to the left between - between M_1 and M_2 the ratio changes drastically until it is almost the inverse of the original ratio.

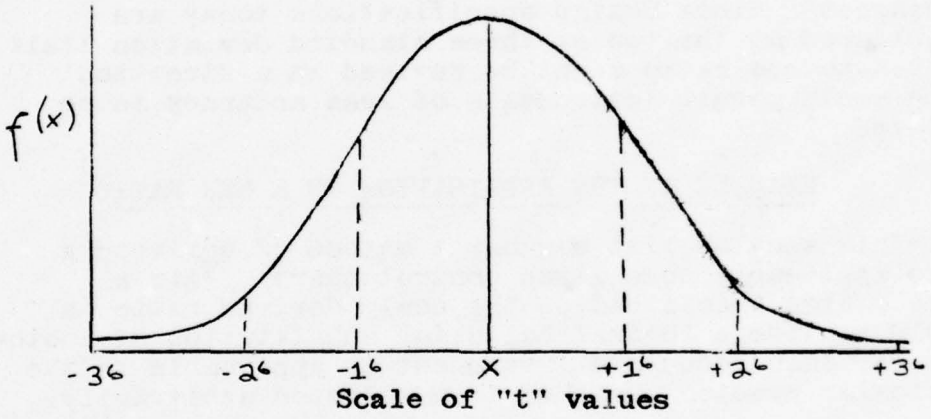
In realistic terms, this means that if our assumptions are true, it may be necessary to shift the emphasis to calibration and standardization techniques etc. - none of which are concerned with the technological construction of the instrument - technological construction of an instrument is related almost entirely to the force/ friction ratio of the receiving or measuring element and the linkage system and accessory units which the element is required to drive. Thus at the higher accuracy instruments at the tightest magnitude of measurement possible, the ten-to-one ratio may possibly be abandoned.

THE "PROBABLE ERROR"

The notion of probable error is now very much obsolescent.[4] However, since the ten-to-one ratio was derived while the "probable error" concept was still popular, some discussion of an approach to a new ratio should be based upon this function.

The understanding of the Normal Probability Curve is rather general. The scale of "t" values (corresponding to standard deviations) is graphically illustrated by figure III.

FIGURE III



Scale of "t" values

The term "probable error" notes + 0.6745σ, and represents a deviation just as likely to be exceeded as not that is, the 50 percent probability that the value has been met.

The probable error always was considered an inaccurate statistical tool. Although it is doubtful that a mathematical basis for the ten-to-one ratio existed, there is strong evidence that if it were supported from statistical methodology, it would have been based upon the probable error rather than the presently usual three-standard deviation concept. For

[4] Facts from Figures - M. J. Moroney, Penguin Books, Baltimore, 1953. P. 114

example, 50 percent of the area under the normal curve falls within the "probable error" area. A plus or minus five percent allowance for exogenous factors (in this case, instrument error) gives a ten-to-one ratio of endogenous to exogenous variables.

This 5 percent allowance at the tails of the distribution is still a very popular one and is reflected in the ± 2 sigma standard deviation, representing approximately 95 percent of the area under the normal curve.

The inaccuracies inherent in a probable error analysis require particularly stringent instrument tolerances. Since design specifications today are established at the two or three standard deviation limit, the ten-to-one ratio might be revised in a direction which would permit instruments of less accuracy to be employed.

EXAMPLE OF THE APPLICATION OF A NEW RATIO

This section will suggest a method of deriving a ratio applicable to a given control chart. With a given design specification the newly derived ratio should provide a logical basis for substitution of another more (or less) accurate instrument as applicable to the particular problem - rather than selected arbitrarily.

A Full-Wave Rectifier was being tested for use in an instrumentation circuit. One hundred fifteen (115) volts 400 c.p.s. alternating current was being put into a black box and specifications required that an output of 150 volts direct current be achieved. Particular attention was directed toward the low temperature range to determine the operation of this unit under severe arctic environment conditions.

At the lower temperatures, the capacitance was reduced and ripple voltage increased (presumably eratically) to a point where it exceeded the 200 millivolts r.m.s. upper control limit of the design specification.

The instrument employed to determine the ripple was an AC Vacuum-Tube voltmeter with a manufacturing accuracy of 3 percent. However, a 2 percent accuracy was obtained through calibration. On the one volt scale this corresponded to ± 20 millivolts.

Assuming that the design tolerance had been set at 200 millivolts, the ratio of design tolerance to instrument accuracy was ten-to-one

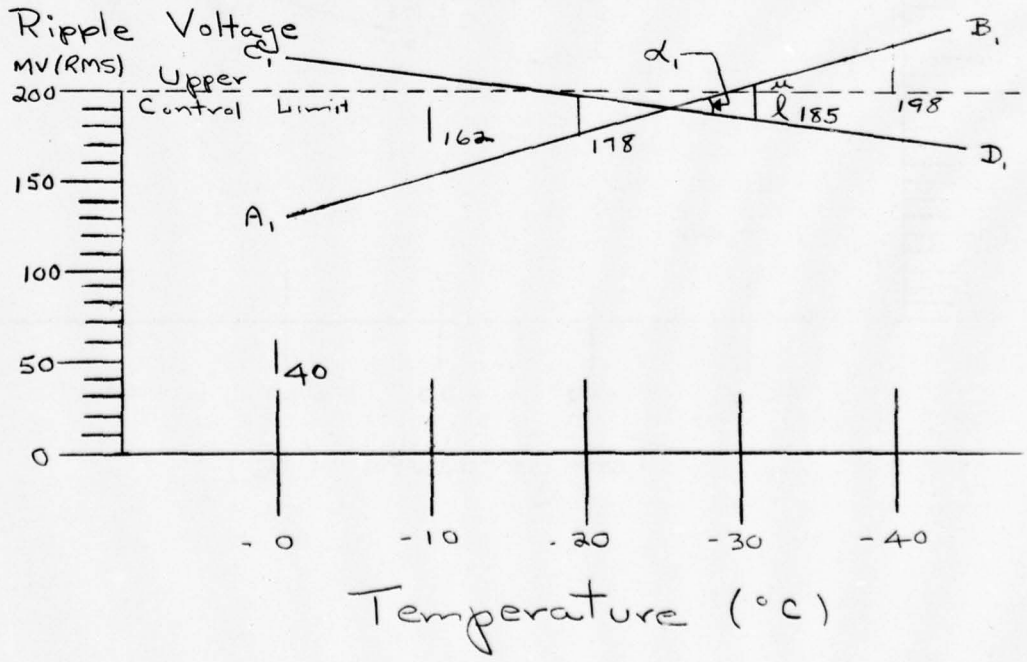
Temperature readings were taken within the range of $0^{o}C$ and -50^{o} and ripple increased from 40 mv until the design tolerance was exceeded at $-30^{o}C$. (See Figure IV). This actual reading at $-30^{o}C$ was 185 mv. Test specifications required that +20 mv be added to the actual reading.

However, it was determined that more accurate instrumentation should be employed so that it might be possible to bring this equipment to within specification limits at $-30^{o}C$.

The use of a 1 percent voltmeter was recommended. The tolerance would be tightened to a + 10 mv, and the rectifier would be within specified tolerance.

The derived ratio, however, is now 20 to 1. The decision to employ this tighter ratio can be justified statistically only by a thorough analysis of the data. New mathematical techniques - presumably based upon the expected variability of the process - must be standardized employed in industry.

FIGURE IV



The process has gone out of control at the -30°C reading due to the wide spread allowed for instrument error. The 200 mv has been used as a control limit and is based solely upon the design tolerance. The ratio of the instrument accuracy to the design tolerance could be based upon the slope between the upper instrument tolerance and the lower instrument tolerance of the previous reading. Angle α_1 is the ratio function.

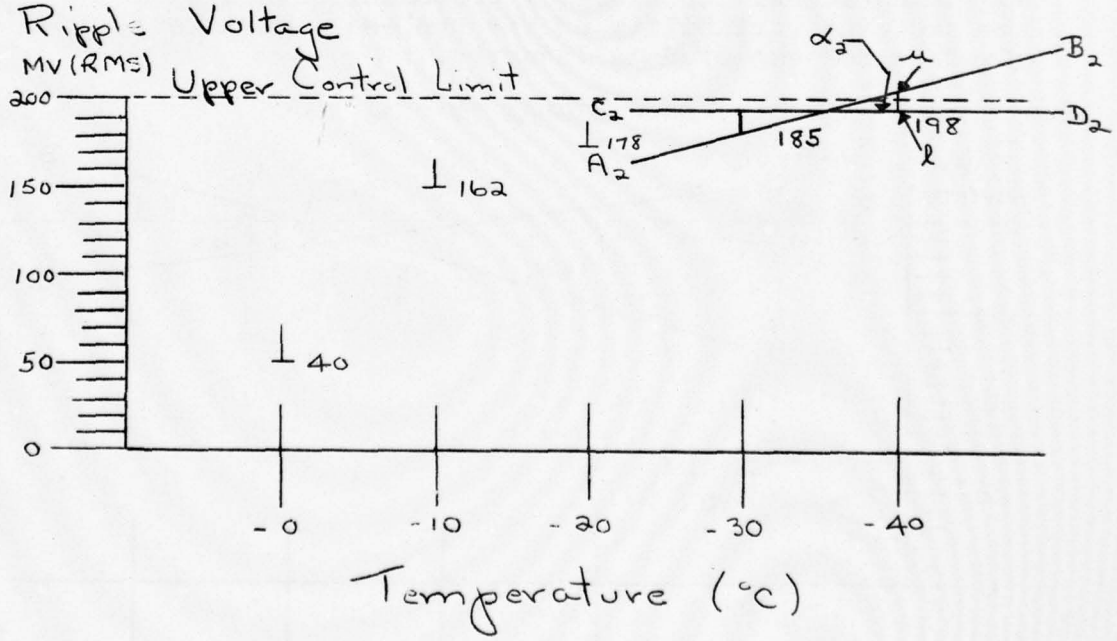
$$f(\alpha_1) = \frac{Ti_1}{Td} < \frac{1}{10} \qquad (1)$$

where Ti_1 is the instrument tolerance

Td is the design tolerance

Assume the same readings taken with the second instrument and an angle α_2 is derived in a similar manner, thus:

FIGURE V



In this case an instrument tolerance of 10 mv (Ti_2) has presented a more accurate picture of the process going out of control. Thus we have the angle (α_2) as a ratio function.

$$f(\alpha_2) = \frac{Ti_2}{Td} < \frac{1}{10} \qquad (2)$$

The design tolerances remain at 200 mv. The instrument tolerance is now to be tightened and it is required to determine to what new ratio we must work.

Where formerly, from equation (1)

$$f(\bar{\alpha}_1) = \frac{Ti_1}{Td} = \frac{1}{10}$$

(The bar denotes known values)

now

$$f(\bar{\alpha}_2) = \frac{Ti_2}{Td}$$

equating (1) + (2)

$$Ti_2 = \frac{f(\bar{\alpha}_2)\bar{Ti}_1}{f(\bar{\alpha}_1)} \qquad (3)$$

The $f(\alpha_1)$ function is derived as follows:

Let the upper tolerance value be designated as "u"; let the lower tolerance value be designated as "l"

Then the angle generated by the slope of the line $\overline{A_1B_1}$ is computed as:

$$\frac{u_4 - l_3}{(4 - 3)} = \text{Tan } \theta_1 \qquad (4)$$

The angle generated by the slope of the line $\overline{C_1D_1}$ is computed as:

$$\frac{u_3 - l_4}{(3-4)} = \text{Tan } \phi_1 \qquad (5)$$

Since α is composed of positive contribtuions from θ & ϕ,

$$\theta_1 - \phi_1 = \alpha_1$$

in the general case:

$$\theta - \phi = \alpha \qquad (6)$$

Similarly, the $f(\alpha_2)$ function is derived from the figure V.

$$\frac{u_5 - l_4}{(5-4)} = \tan \theta_2$$

$$\frac{u_4 - l_5}{(4-5)} = \tan \phi_2$$

from equation (6)

$$\theta_2 - \phi_2 = \alpha_2$$

This process is repeated until the entire previous history of the phenomenon is incorporated into a new instrument accuracy requirement. The purpose of this derivation is to present a value (α) which could be used in a new ratio. The function incorporates both a factor related to instrument accuracy and variability of the process. For example, where the process is not excessively variable, tight instrument requirements are justified. Therefore there exist some justification for a tighter ratio; 20 to 1.

CONCLUSION

The purpose of presenting these statistical tools has been to suggest an approach to the derivation of a new standard. Obviously, there could be many other methods of deriving new ratios. It has been demonstrated

that the previous standard is inapplicable. However, any new standard, by definition, requires the concerted approval of those industrial operations most directly affected.

The need for this derivation has been clearly expressed to the author during consultation with various industrial organizations. The approaches proposed above are neither complete nor mathematically rigorous and no attempt is made to make them so - this being beyond the scope of the report. It remains for the statisticians of the industrial organizations to agree upon the derivation of a set of ratio standards satisfactory to industry as a whole.

The initial step in such a program is a complete compilation of literature available both here and abroad related to instrument accuracy. The enclosed biblioabstract is intended as an initial step in the performance of this function.

STATISTICAL DEVELOPMENTS IN LIFE TESTING[a]

Benjamin Epstein
Department of Statistics
Wayne State University
Detroit, Michigan

Summary

In this paper we describe recently developed statistical methods for analyzing data arising from life tests and for designing life tests. Advantage is taken of the time ordered nature of life test data to shorten substantially the time required to reach a decision. Most of the results have been obtained under the assumption of an exponential distribution of life. Replacement, non replacement, sequential, non sequential, and truncated procedures are described. Some useful tables are given at the end of the paper.

It is a characteristic feature of most life and fatigue tests that they give rise to ordered observations. If, for example, twenty radio tubes are placed on life test and t_i denotes the time when the ith tube fails, the data occur in such a way that $t_1 \leq t_2 \leq \cdots \leq t_{20}$. Exactly the same kind of ordered situation will occur whether the problem under consideration deals with the life of electric bulbs, the life of electronic components, the life of ball bearings, or the length of life of human beings after they are treated for a disease. The examples we have just given all involved ordering in time. This need not necessarily be the case. If we are interested in destructive test situations involving such things as the current needed to blow a fuse, the voltage needed to break down a condenser, the force needed to rupture a physical material, then we can often arrange to test in such a way that every item in the sample is subjected to precisely the same stimulus (current, voltage, stress). If this is done, then clearly the weakest item will be observed to fail first, the second weakest next, etc. In the present paper we discuss almost exclusively situations in which it is the time to failure that is the important random variable, and therefore we shall use the language of time throughout the paper. It should be emphasized, however, that there will be some practical problems which do not involve time, but for

[a] The preparation of this paper was supported in part by the Office of Naval Research.

which some of the ideas discussed in this paper are quite relevant.

Put in general terms, we test n items drawn at random from some population and the data become available in such a way that the smallest observation comes first, the second smallest second, ..., and finally the largest observation last. Clearly we can, if we choose, discontinue experimentation long before all n items have failed. In particular we may decide to terminate the experiment as soon as we have the first r ($\leq n$) failures, or we may decide to stop at some preassigned truncation time T_o, or we may adopt a sequential procedure permitting us to stop as soon as certain conditions are met. In all of these cases our primary concern is the development of statistical procedures which, by taking advantage of the fact that data become available in order, will enable the experimenter to reach a decision in a shorter time or with fewer observations, than would be possible if data did not become available in a time ordered way.

II. Preliminary remarks on the exponential distribution.

In this paper virtually all results will be obtained under the assumption that the length of life X has an exponential distribution described by the probability density function (henceforth abbreviated as p. d. f.) $f(x;\theta)$ of the form

$$(1) \qquad f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x \geq 0, \quad \theta > 0$$
$$= 0, \text{ elsewhere.}$$

A partial justification for this assumption has been discussed in some detail by the author (see ref(1)) and several relevant references are given in that paper. Quite recently further evidence of an empirical nature can be found in a series of ARINC monographs. We are well aware of the fact that many life distributions are not adequately described by equation (1). However, we feel that an understanding of the theory in the exponential case is essential if we are to treat more general situations. In fact, in some cases, the solution for a p. d. f. which is not of the form (1) can be readily obtained by making trivial modifications of the results in the exponential case. We intend to discuss this question in detail in another paper.

Returning then to the p. d. f. (1) we state some results which are discussed in detail in a paper by Epstein and Sobel (see ref.(3)). The first result is as follows: Let n items

be drawn at random from a distribution whose p. d. f. is given by (1) and placed on life test. Let the observations become available in order, i.e., $x_{1,n} \leq x_{2,n} \leq \cdots \leq x_{r,n} \leq \cdots \leq x_{n,n}$ where by $x_{i,n}$ is meant the time when the i'th failure occurs. Suppose that experimentation is discontinued as soon as the r'th item fails (r is preassigned), then it can be shown that the maximum likelihood estimate of the mean life[b] θ is given by $\hat{\theta}_{r,n}$ where

$$(2) \qquad \hat{\theta}_{r,n} = \frac{x_{1,n} + x_{2,n} + \ldots + x_{r,n} + (n-r)\, x_{r,n}}{r}.$$

In words we add up the total number of hours lived by all items, those that failed and those which did not fail, and divide by the number of failures. The estimate $\hat{\theta}_{r,n}$ is "best" in the sense that in addition to being maximum likelihood, it is also unbiased, minimum variance, efficient, and sufficient. The p. d. f. of $\hat{\theta}_{r,n}$ is given by

$$(3) \qquad f_r(y) = \frac{1}{(r-1)!} (r/\theta)^r \; y^{r-1} \; e^{-ry/\theta}, \quad y > 0$$
$$= 0, \text{ elsewhere}$$

and $2r\hat{\theta}_{r,n}/\theta$ is distributed as chi-square with $2r$ degrees of freedom (which we denote as $\chi^2(2r)$).

In the preceding paragraph we have been concerned with the non-replacement situation where one does not replace failed items at once by new items drawn from the underlying p. d. f. (1). In the replacement case (where one immediately replaces a failed item by a new one) it can be shown that the maximum likelihood estimate of the mean life θ is given by

$$(4) \qquad \hat{\theta}_{r,n} = n\, x_{r,n}/r,$$

where by $x_{r,n}$ is meant the total time (measured from the beginning of the life test) to observe the r'th failure and where the sample size n is maintained throughout the life test. It should be remarked that $nx_{r,n}$ is the total number

[b] θ is the mean life since $E(X) = \int_0^\infty x \frac{1}{\theta} e^{-x/\theta}\, dx = \theta.$

of hours lived by all items on test since

$$(5) \quad nx_{r,n} = nx_{1,n} + n(x_{2,n} - x_{1,n}) + n(x_{3,n} - x_{2,n}) + \ldots + n(x_{r,n} - x_{r-1,n}).$$

On the righthand side of (5), $nx_{1,n}$ is the number of hours lived by all items up to the time the first failure occurred, and $n(x_{i,n} - x_{i-1,n})$ is the number of hours lived by all items between the times of occurrence of the (i-1)st failure and i'th failure. The estimate (4) in the replacement case has precisely the same distribution and the same optimum properties as does the estimate (2) in the non-replacement case. In fact if we let $T_{r,n}$ be the total number of hours lived by all items whether they failed or not, up to the time when the r'th failure occurred, one can write both (2) and (4) as

$$(6) \quad \hat{\theta}_{r,n} = T_{r,n}/r$$

where

$$T_{r,n} = x_{1,n} + x_{2,n} + \ldots + x_{r-1,n} + (n - r + 1)\, x_{r,n}$$

in the non-replacement case and where

$$T_{r,n} = nx_{r,n}$$

in the replacement case. In either case, $2T_{r,n}/\theta$ is distributed as $\chi^2(2r)$.

An interesting and important feature of the distribution of $\hat{\theta}_{r,n}$ in either the replacement or non-replacement case is its independence of n. It therefore follows that no matter what n is a $100(1 - \alpha)$ percent confidence interval for the true but unknown mean life θ based on a test terminated after one has observed the first r out of n failures is given by

$$(7) \quad \left(\frac{2r\hat{\theta}}{\chi^2_{\frac{\alpha}{2}}(2r)}, \frac{2r\hat{\theta}}{\chi^2_{1-\frac{\alpha}{2}}(2r)} \right) = \left(\frac{2T_r}{\chi^2_{\frac{\alpha}{2}}(2r)}, \frac{2T_r}{\chi^2_{1-\frac{\alpha}{2}}(2r)} \right)$$

where we define the constant χ^2_γ (2k) by the equation

$$(8) \qquad \Pr\left(\chi^2(2k) > \chi^2_\gamma(2k)\right) = \gamma.$$

Similarly, suppose we want to find a test procedure which will give a prescribed operating characteristic curve (henceforth abbreviated as O. C. curve). Put in statistical terms[c] we want to test the hypothesis $H_o : \theta = \theta_o$ against the alternative $H_1 : \theta = \theta_1 < \theta_o$ subject to the conditions that for $\theta = \theta_o$, $L(\theta_o) = \Pr$(accepting $\theta = \theta_o \mid \theta_o$ is true) $= 1 - \alpha$ and for $\theta = \theta_1$, $L(\theta_1) = \Pr$ (accepting $\theta = \theta_o \mid \theta_1$ is true) $\leq \beta$. It is shown in our paper (3) that the region of acceptance for $\theta = \theta_o$ must be of the form

$$(9) \qquad \hat{\theta}_{r,n} > C = \theta_o \, \chi^2_{1-\alpha}(2r)/2r,$$

where the O. C. curve based on this region of acceptance must be independent of n, since the distribution of $\hat{\theta}_{r,n}$ depends only on r. The appropriate values of r (and hence C) for certain values of α, β, and θ_o/θ_1 are given in Table 1. For values of α, β, and θ_o/θ_1 not given in the table, the appropriate r to use is the smallest integer r such that $\chi^2_{1-\alpha}(2r) \Big/ \chi^2_\beta(2r) \geq \theta_1/\theta_o$.

In the test procedure $\hat{\theta}_{r,n} > C$, the sample size n is at our disposal. The effect of increasing n is to shorten the time needed on the average to reach a decision and thus if we happen to be in a situation where the items being tested are cheap but where time is very valuable, we may well prefer a test of the form $\hat{\theta}_{r,n} > C$ to one which is of the form $\hat{\theta}_{r,r} > C$. These two procedures have exactly the same O. C. curve and our only reason for preferring a rule of action based on the first r failures out of n items tested to one based on failing all r out of r items is that the first rule of action will take a shorter time on the average. Thus, for example,

---

(c) θ_o is some acceptable (high) mean life, θ_1 is some unacceptable (low) mean life, α is the producer's risk and β is the consumer's risk.

a test procedure which involves stopping an experiment after the first of two items on test has failed will lead to rules of action whose O. C. curve is precisely the same as that found by placing one item on test and waiting until it fails. However, the expected length of time in the first procedure is only one half that in the second procedure. Consequently, if the time saved outweighs the loss due to testing two items rather than one, we would prefer the first procedure.

Let $E(X_{r,n})$ be the expected length of time needed to observe the first r failures out of n items placed in test, and let $E(X_{r,r})$ be the expected length of time needed to observe all r items to fail, if r items are placed on test, then the ratio

$$(10) \qquad \alpha_{r,n} = E(X_{r,n})/E(X_{r,r})$$

is a measure of the expected saving in time due to using the first procedure as compared with the second procedure. In table 2 we give the values of this ratio for selected small values of r and n, in the non-replacement case. This table shows, that if "time is money", procedures which terminate before the whole sample is observed may be very advantageous. In evaluating (10) the following formulae are useful:

$$(11) \qquad E(X_{r,n}) = \theta\left(\frac{1}{n} + \frac{1}{n-1} + \cdots \frac{1}{n-r+1}\right) = \theta \sum_{j=1}^{r} 1/(n-j+1).$$

in the case where failed items are not replaced and

$$(12) \qquad E(X_{r,n}) = r\theta/n$$

in the case where failed items are replaced at once by new items drawn from the p. d. f. (1)

III Truncated life tests

It is frequently necessary on practical grounds to terminate a life test by a preassigned time T_o. This leads to truncated tests in which it is decided in advance that the life test will be terminated at min $(X_{r_o,n};\ T_o)$ where $X_{r_o,n}$ is the time at which the r_o'th failure occurs and T_o is the truncation time beyond which the life test

will not be allowed to run. If the life test is terminated at $X_{r_o,n}$ (i.e., r_o failures occur before time T_o) then the action taken will be to reject. If the experiment is terminated at time T_o (i.e., the r_o'th occurs after time T_o) then the action in terms of "hypothesis" testing is acceptance. In a paper by Epstein (see ref. (4)) one can find details concerning such test procedures for both the replacement and non-replacement cases. These test procedures are characterized by three functions $E_\theta(r)$, $E_\theta(T)$, and $L(\theta)$, the expected number of observations to reach a decision, the expected waiting time to reach a decision, and the probability of accepting respectively, if θ is the true value. The formulae are given below.

In the non-replacement case

$$(13)\quad E_\theta(r) = np_\theta \left[\sum_{k=0}^{r_o-2} b(k;n-1,p_\theta) \right] + r_o \left[1 - \sum_{k=0}^{r_o-1} b(k;n,p_\theta) \right]$$

where

$$p_\theta = 1 - e^{-T_o/\theta} \quad \text{and} \quad b(k;n,p_\theta) = \binom{n}{k} p_\theta^{\,k} (1-p_\theta)^{n-k}$$

The probability distribution of r is given by

$$(14)\quad \Pr(r = k \mid \theta) = b(k;n,p_\theta),\ k = 0, 1, 2, \ldots, r_o - 1$$

and

$$(14')\quad \Pr(r = r_o \mid \theta) = 1 - \sum_{k=0}^{r_o-1} \Pr(r = k \mid \theta).$$

Further one has

$$(15)\quad E_\theta(T) = \sum_{k=1}^{r_o} \Pr(r = k \mid \theta)\ E_\theta(X_{k,n})$$

where $E_\theta(X_{k,n})$ can be found from (11), and

$$(16)\quad L(\theta) = \sum_{k=0}^{r_o - 1} \Pr(r = k \mid \theta).$$

In the replacement case the probability distribution of r is given by

$$(17)\quad \Pr(r = k \mid \theta) = p(k;\ \lambda_\theta),\ k = 0,\ 1,\ 2,\ldots,\ r_o - 1$$

and

$$(17')\quad \Pr(r = r_o \mid \theta) = 1 - \sum_{k=0}^{r_o - 1} p(k;\ \lambda_\theta).$$

In (17) and (17'), $\lambda_\theta = nT_o/\theta$ and $p(k;\ \lambda_\theta) = \lambda_\theta^{\,k} \exp - (\lambda_\theta)/k!$.

Further one has

$$(18)\quad E_\theta(r) = \lambda_\theta \sum_{k=0}^{r_o - 2} p(k;\lambda_\theta) + r_o \left[1 - \sum_{k=0}^{r_o - 1} p(k;\lambda_\theta)\right]$$

$$(19)\quad E_\theta(T) = \theta E_\theta(r)/n$$

and

$$(20)\quad L(\theta) = \sum_{k=0}^{r_o - 1} p(k;\ \lambda_\theta).$$

We have just given formulae for the O. C. curve, the expected waiting time, and expected number of items failed in the course of reaching a decision for any preassigned n, T_o, r_o. We now give a formula for finding the appropriate truncated test (that is, for finding r_o and n) when the truncation time T_o is preassigned and the O. C. curve is required (for preassigned type I error, α, and type II error, β) to be such that $L(\theta_o)=1-\alpha$ and $L(\theta_1)\leq\beta$. It is proved in the paper referred to in the first paragraph of this section that for both the replacement case and the non-replacement case the appropriate r_o is precisely the same as the r_o used in tests of the form (9). Hence Table 1 can be used. As for the appropriate value of n one should choose

$$n = \left[\theta_o \, \chi^2_{1-\alpha}(2r_o)/2T_o\right] \tag{21}$$

where $[x]$ means the greatest integer $\leq x$, in the replacement case.

In the non-replacement situation a good approximate value of n, in case θ_o/T_o is substantially more than one (say ≥ 3), is given by

$$n = \left[r_o/(1 - e^{-T_o/C})\right] \tag{22}$$

where $$C = \theta_o \, \chi^2_{1-\alpha}(2r_o)/2r_o \, .$$

IV. Sequential Life Tests

One can make substantial improvements on the procedures described in sections II and III by following a sequential procedure. It is shown in a paper by Epstein and Sobel (see ref.(5)) that the sequential probability ratio test of A. Wald can be applied to life testing. It is very interesting that decisions can now be made continuously in time. At each moment t, one can decide either to accept, to reject, or to

continue the life test. If we are, as before, testing H_o: $\theta = \theta_o$ against H_1: $\theta = \theta_1$ ($\theta_o > \theta_1$) with Type I error = α and Type II error = β, then the decision as time unfolds depends on

$$(23) \qquad B < (\theta_o/\theta_1)^r \exp - \left\{ (1/\theta_1 - 1/\theta_o)V(t) \right\} < A$$

where A and B can for all practical purposes be taken as

$$(24) \qquad A = (1 - \beta)/\alpha \quad \text{and} \quad B = \beta/(1 - \alpha).$$

In (23), r is the number of failures observed by time t. The decision to continue experimentation is made as long as the inequality (23) holds. As soon as (23) is violated, one accepts H_o if the function of t in (23) is $< \beta$, and one rejects H_o (accepts H_1) if the function of t in (23) is $> A$.

In (23) V(t) is a statistic which equals the total number of hours lived by all items, failed and unfailed, up to time t. In the replacement case

$$(25) \qquad V(t) = nt,$$

while in the non-replacement case[d]

$$(26) \qquad V(t) = \sum_{i=1}^{r} (n - i + 1)(x_i - x_{i-1}) + (n - r)(t - x_r)$$

$$= \sum_{i=1}^{r} x_i + (n - r)(t - x_r),$$

It is convenient to write (23) as

$$(27) \qquad -h_1 + rs < V(t) < h_o + rs,$$

where h_o, h_1, and s are positive constants given by

$$(28) \qquad h_o = \frac{-\log B}{1/\theta_1 - 1/\theta_o}, \quad h_1 = \frac{\log A}{1/\theta_1 - 1/\theta_o}, \quad s = \frac{\log (\theta_o/\theta_1)}{1/\theta_1 - 1/\theta_o}.$$

[d] It should be remarked that in the non-replacement case a special problem arises if all n items fail without reaching a decision. This eventually can be taken care of in various ways.

It is shown in our paper referred to in the first paragraph how formula (27) enables one to carry out the sequential procedure graphically.

The O. C. curve, that is, the probability of accepting H_o when θ is the time parameter value, is given approximately by a pair of parametric equations

$$(29) \qquad L(\theta) = \frac{A^h - 1}{A^h - B^h}, \qquad \theta = \frac{(\theta_o/\theta_1)^h - 1}{h(1/\theta_1 - 1/\theta_o)},$$

by letting the parameter h run through all real values.

The values of $L(\theta)$ at the five points $\theta = 0,\ \theta_1,\ s,\ \theta_o,\ \infty$ enable one to sketch the entire curve. These values are respectively 0, β, $\log A/(\log A - \log B)$, $1 - \alpha$, and 1.

$E_\theta(r)$, the expected number of observations required to reach a decision when θ is the mean life is given by

$$(30) \qquad E_\theta(r) \sim \begin{cases} \dfrac{h_1 - L(\theta)\,(h_o + h_1)}{s - \theta}, & \theta \neq s \\ \dfrac{h_o\, h_1}{s^2}, & \theta = s \end{cases}$$

If we let $k = \theta_o/\theta_1$, the approximate values of $E_\theta(r)$ become particularly simple when $\theta = \theta_1$, s, or θ_o. They are

$$(31) \qquad E_{\theta_1}(r) \sim \left[\beta \log B + (1 - \beta) \log A\right] \Big/ \left[\log k - (k - 1)/k\right]$$

$$E_s(r) \sim -\log A \log B \,/\, (\log k)^2,$$

$$E_{\theta_o}(r) \sim \left[(1 - \alpha) \log B + \alpha \log A\right] \Big/ \left[\log k - (k - 1)\right].$$

In Table 3, we give $E_\theta(r)$ for five values of θ $(0,\ \theta_1,\ s,\ \theta_o,\ \infty)$ for four values of k(3/2, 2, 5/2, 3), and for the four number pairs (α, β) which can be made with the numbers .01 and .05.

It can be shown that $E_\theta(t)$, the expected waiting time to reach a decision is given by the formula

$$(32) \qquad E_\theta(t) = E_\theta(r)\ \theta/n$$

in the replacement case. In the non-replacement case,

$$(33) \qquad E_\theta(t) = \sum_{k=1}^{n} \Pr(r = k \mid \theta)\ E_\theta\ (\chi_{k,n})$$

where $E_\theta\ (\chi_{k,n}')$ can be found from (11). A good approximation for $E_\theta(t)$ is given by

$$(34) \qquad E_\theta(t) \sim \theta \log \left(\frac{n}{n - E_\theta\ (r)} \right) .$$

The derivations of all formulae in this section can be found in the reference cited in the first paragraph.

V. Conclusion

We have not attempted in this paper to cover all of the papers which have been published by a number of writers including the author in the field of life testing. We have selected essentially three papers (see ref. (3),(4),(5)) which give some of the results which we consider to be most fundamental. A careful reading of these papers give a good introduction to the statistical methodology involved in life testing. These papers also contain many numerical illustrations which will be of substantial help in seeing how one applies the theory to the design and analysis of life tests.

References

(1) Benjamin Epstein, "Statistical Problems in Life Testing". Proceedings of the Seventh Annual Convention of the American Society for Quality Control, pp. 385 - 398, 1953.

(2) Aeronautical Radio Inc., "Investigation of Electronic Equipment Reliability as Affected by Electron Tubes", Inter-base Report No. 1, March 15, 1955.

(3) B. Epstein and M. Sobel, "Life Testing", Journal of the American Statistical Association 48, 485 - 502, 1953.

(4) B. Epstein, "Truncated Life Tests in the Exponential Case", Annals of Mathematical Statistics 25, 555 - 564, 1954.

(5) B. Epstein and M. Sobel, "Sequential Life Tests in the Exponential Case", Annals of Mathematical Statistics 26, 82 - 93, 1955.

Table 1

Values of r (upper numbers) and of $\chi^2_{1-\alpha}(2r)/2$ (lower numbers) such that the test based on using $\hat{\theta}_{r,n} > C = \theta_o \chi^2_{1-\alpha}(2r)/2r$ as acceptance region for $\theta = \theta_o$ will have $L(\theta_o) = 1 - \alpha$ and $L(\theta_o) \leq \beta$.

θ_o/θ_1	$\alpha = .01$			$\alpha = .05$			$\alpha = .10$		
	$\beta = .01$	$\beta = .05$	$\beta = .10$	$\beta = .01$	$\beta = .05$	$\beta = .10$	$\beta = .01$	$\beta = .05$	$\beta = .10$
3/2	136 110.4	101 79.1	83 63.3	95 79.6	67 54.1	55 43.4	77 66.0	52 43.0	41 33.0
2	46 31.7	35 22.7	30 18.7	33 24.2	23 15.7	19 12.4	26 19.7	18 12.8	15 10.3
5/2	27 16.4	21 11.8	18 9.62	19 12.4	14 8.46	11 6.17	15 10.3	11 7.02	9 5.43
3	19 10.3	15 7.48	13 6.10	13 7.69	10 5.43	8 3.98	11 7.02	8 4.66	6 3.15
4	12 5.43	10 4.13	9 3.51	9 4.70	7 3.29	6 2.61	7 3.90	5 2.43	4 1.75
5	9 3.51	8 2.91	7 2.33	7 3.29	5 1.97	4 1.37	5 2.43	4 1.75	3 1.10
10	5 1.28	4 .823	4 .823	4 1.37	3 .818	3 .818	3 1.10	2 .532	2 .532

Table 2

Ratio of the Expected Waiting Time to Observe the r'th Failure in Samples of Size n and r respectively

$$E(X_{r,n}) \,/\, E(X_{r,r}) = \alpha_{r,n}$$

r \ n	1	2	3	4	5	10	15	20
1	1	.50	.33	.25	.20	.10	.067	.050
2	-	1	.56	.39	.30	.14	.092	.068
3	-	-	1	.59	.43	.18	.12	.087
4	-	-	-	1	.62	.23	.14	.104
5	-	-	-	-	1	.28	.18	.125
10	-	-	-	-	-	1	.35	.23

Table 3

Approximate values of $E_\theta(r)$ for sequential tests for various values of $k = \theta_0/\theta_1$ and α, β.

$k = \theta_0/\theta_1$		3/2		2		5/2		3	
α		.01	.05	.01	.05	.01	.05	.01	.05
θ	β								
0	.01	11	7	7	4	5	3	4	3
	.05	11	7	7	4	5	3	4	3
θ_1	.01	62.4	40.3	23.3	15.1	14.2	9.20	10.4	6.74
	.05	60.4	36.7	22.6	13.7	13.8	8.38	10.1	6.14
	.01	128	82.7	43.9	28.3	25.1	16.2	17.5	11.3
	.05	82.7	52.7	28.3	18.0	16.2	10.3	11.3	7.18
θ_0	.01	47.6	44.2	14.7	13.6	7.71	7.16	5.00	4.63
	.05	30.8	28.0	9.48	8.64	4.99	4.54	3.23	2.94
∞	any	0	0	0	0	0	0	0	0

ANALYSIS OF VARIANCE MODELS WITH ENGINEERING APPLICATIONS

By Mary D. Lum
Wright Air Development Center

I. PURPOSE

As the title indicates, I am going to discuss "Analysis of Variance Models with Engineering Applications". There are two main points which I propose to cover. The first is to emphasize to you the difference between fixed and random factors in an experiment, and their influence upon statistical tests and inferences.

The second main point is to describe (by means of examples) an analysis of variance model more general than the factorial or pure hierarchal types. It is a type of nesting design which I shall call "partially hierarchal".

For those of you who are interested in obtaining references to the Analysis of Variance procedure and its applications, here are a few reports which would make a good start on the subject:

(1) WADC TR 56-20, "An Elementary Approach to the Analysis of Variance" by Rider, Harter, and Lum. ASTIA No. AD93394. (Contains a large bibliography to other papers and books)

(2) WADC TR 55-33 "Partially Hierarchal Models in the Analysis of Variance" by Harter and Lum. ASTIA No. AD75480.

(3) WADC TR 53-23 "Tests by the Analysis of Variance" by Mentzer. ASTIA No. AD14028

(4) Eisenhart - The assumptions underlying the analysis of variance. Biometrics 3, (1947) pp 1-21. (Model I, Model II).

(5) Cochran - Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics 3 (1947) pp 22-38.

I will define FACTOR[1], as a suspected source or cause of variation taken into account by the experiment, LEVEL as one

[1] There are two types of factors:
(1) "main" factors
(2) interactions of two or more "main" factors

condition of a factor, and EFFECT as a numerical value associated with a level of a factor. A fixed factor (F) is distinguishable where the levels of F constitute a population and a random factor (f) is distinguishable where levels of f represent a random sample from an infinite population of such levels.

With regard to the first point (on the differentiation between fixed and random factors) it is possible on occasion to arrive at essentially the same statistical result regardless of whether the factor is fixed or random. However, in general this is not the case. To emphasize this point, I am going to present an experiment in which "different" answers are obtained depending on the type of factor-fixed or random. If you are making inferences about a particular set of experimental levels (fixed factor) in general you make different statistical tests than if you wish to generalize by making inferences about an infinite population from which your experimental levels were assumed to be chosen at random (random factor). The statistical conclusions based on these tests may then be "different", in a narrow sense. Taken from a broader viewpoint they are really just different aspects or facets of the same physical situation and when used properly do not lead to inconsistent results.

Misinterpretation of this difference between fixed and random factors has led to misunderstandings between statisticians and subject matter specialists, such as the engineer or chemist. In fact this could very well be one more of those touchy points which have given credence to such disparaging remarks as "there are liars, damned liars, and then statisticians". Furthermore, there are those unacquainted with statistical methods (mathematicians included) who believe the statistician must be akin to a magician, for he seems utterly capable of drawing any conclusions he wants from a given set of data. From a superficial point of view this is partly right. It is possible to draw different statistical conclusions from the same set of data depending on the particular statistical tests made. However, if one delves further into a more serious consideration of the nature of the data, how they were obtained, what inferences are possible, and what questions one desires to answer, one finds that the statistical tests to be made depend uniquely on these considerations.

If you take in some regular manner a set of levels of a factor (e.g. temperature at 25°C, 50°C, 75°C, 100°C) for an experiment and can NOT by any stretch of the imagination even remotely regard it as a random sample from some infinite population then naturally it does not make sense to consider it a random factor. It is ipso facto a fixed factor. On the other hand, experimental units (considered as levels of a factor) while not necessarily chosen randomly (in the statistical sense) may be picked in a haphazard manner. If they exhibit no evidence of any regular pattern of relationship to one another, then there is little argument against considering them to be levels of a random factor in order to enable one to generalize. However, one must be extremely careful concerning the type of population to which one is generalizing.

Thus whether considering a factor fixed has more meaning than considering it random, or vice versa depends on (1) what type of statistical inference you wish to make and (2) how the data are taken (this constrains the type of inference you are allowed to make).

II. BACKGROUND

For the benefit of those who are not acquainted with it, I would like to backtrack a little to give a definition and a brief elementary description of what the analysis of variance procedure is. The analysis of variance is a statistical technique which separates the total variance in a set of data into parts, each representing a linear combination of the variances of different factors. A description of the significance of the factor F is depicted as follows:

Let A estimate a linear combination of variances which includes the variance of F

Let B estimate the same linear combination (as estimated by A) with the deletion of the term involving the variance of F.

Let $F = \frac{A}{B}$

If $A \gg B$, i.e. if $F > F_{\alpha} > 1$

then the factor F is "significant" at the α level. (Equivalently, the set of effects corresponding to the levels of F can be said to "vary significantly" at the α level)

If A is of sufficiently greater magnitude than B, then one can conclude that the factor under consideration has contributed significantly to the variation. Thus, in this manner the Analysis of Variance facilitates determining whether the factor under consideration has significantly influenced the variation in the data.

What the linear combinations consist of will depend on whether the factors are fixed or random. The linear combinations in turn determine what F-tests are to be made.

Now consider the following experiment on Human Engineering. The experiment was performed by the Acceleration Section, Biophysics Branch of the Aero Medical Laboratory at Wright Air Development Center.

III. CENTRIFUGE EXPERIMENT

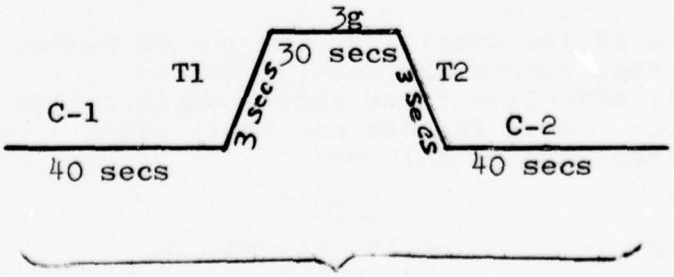
The purpose was to compare the effects of a protected condition (pressure suit inflated) vs an unprotected condition (pressure suit deflated) on human performance in tracking a target while subjected to a 3g force. A measure of performance is the time on target. The inflated pressure consisted of an additional 1-1/2 to 2 lbs/sq. in. applied at abdomen, thighs and calves of legs for maintaining better equilibrium. The experiment was performed in a centrifuge.

Sources of Variation

Other main factors that were suspected of affecting the test results were:

(1) Acceleration conditions (A): Control 1, (Standing still), Transition 1, 3G, Transition 2, Control 2, (standing still).

A = fixed factor



This is called a "run"

Rest Period:

3 minutes between runs

The transition data have not been analyzed.

(2) Amount of Learning: Fatigue

In order to evaluate this effect, 10 runs were made with 3 minute rests in between. A physiologist who consulted on this problem considered this a sufficient rest period for the human body to return to normal activity (runs considered random).

(3) Personal reaction to the stress (Subjective feeling) individual effect. (S)

The group of 6 subjects consisted of volunteers. It will develop later that the experience of the subjects had an important effect on the test results. The group consisted of the following:

Rank of Experience	Code letter	Subjects
3	M	Airman, maintenance
1	E	Physiologist, rated pilot
5	Sa	Medical Doctor
4	L	Physiologist
6	Si	Psychiatrist
2	B	Sergeant, maintenance

All the subjects in the experiment were of medium build. (obese people make poor subjects as they may black out sooner)

Forces to be considered are:

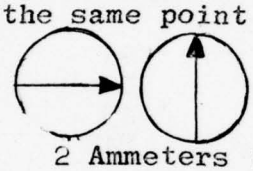
(1) Positive g (Blackout) --- from head to toe
(2) Negative g (Redout) --- from toe to head
(3) Transverse (tangential or centripetal) -- from chest to back

Because of the nature of the centrifuge and the extremely short duration under stress, forces (2) and (3) may be neglected. Thus the only effective force under consideration is the "positive" 3g force. Main factors and their types in the centrifuge experiment are as follows:

Main Factors	Type	Levels
Protection	Fixed	with, without (2)
Acceleration	Fixed	Control 1, 3g, Control 2(3)
Subjects	Fixed or Random	S_1, S_2, ..., S_6 (6)
Runs	Random	R_1, R_2, ..., R_{10} (10)

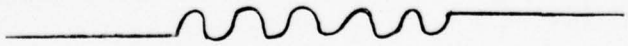
Description of the Experiment

Subject donned G-suit and was prepared for bioelectric measurements. Subject then sat in a prototype of a pilot's seat located inside an enclosed cabin. This cabin is at one end of a long arm, the other end being fixed (the centrifuge). About three feet in front of him at eye-level are located two ammeters, each about 4 inches in diameter with 2 sine slow wave oscillators (approximately .05 cycles/sec) driving the two needles. The oscillators are at slightly different frequencies in order to keep the needles from crossing the same point at the same time.



2 Ammeters (in milliamperes)

The pointers are hinged as illustrated and move on scales marked from -100 to +100 (a radial distance of ± 1 inch). A stick simulating the pilot's control stick is located directly in front of the subject and between his legs. Movement of the stick controls the movements of the two pointers. The task assigned to the subject is to keep both pointers on a spot corresponding to the center of the dials by movement of his stick. The two pointers are actuated away from this point by electrical means. The subject is considered to be "on target" if neither pointer is more than ± 25 units away from the center spot (about 1/4"). An oscillograph attached to a pen writeout records his performance. If the pointers are within ± 25 units, i.e. "on target", the pen writes a straight line; if the pointers are not within this limit, "not on target", a 60 cycle frequency is actuated:



The subject had two trials, one for each pressure condition. A trial consisted of 10 runs, each run being performed as I described before. At the 3G condition the centrifuge is rotating at such a speed as to produce a centrifugal force equivalent to 3G, (96.6 ft/sec^2) on the subject.

It was assumed that the order of presentation of the pressure conditions made no difference on the performance in a trial. However, as an extra precaution the order of presentation was randomly selected. Some subjects performed

first with the pressure "on"; others performed first with the pressure "off". Between one and three weeks elapsed between the first trial and the second.

A mathematical model for this experiment is given by the following:

FACTORS

		(f=Fixed) (r=random)
Protection (f)	PA	
Acceleration (f)	PS	
Subjects (f or r)	AS	
Runs within PAS (r)	PAS	

MATHEMATICAL MODEL

$$\text{Time on target} = M+P_i+A_j+S_k+(PA)_{ij}+(PS)_{ik}+(AS)_{jk}+(PAS)_{ijk}+r_{ijkl}$$

ASSUMPTIONS

M = constant

$P_1 = -P_2$ = constant; A_1, A_2, A_3 are constants $\sum_j A_j = 0$

$(PA)_{ij}$ are constants, $\sum_i (PA)_{ij} = \sum_j (PA)_{ij} = 0$

$r_{ijkl} = N(0, \sigma_r^2)$

If S = FIXED

S_k are constants, $\sum_k S_k = 0$

$(PS)_{ik}$ are constants, $\sum_i (PS)_{ik} = \sum_k (PS)_{ik} = 0$

$(AS)_{jk}$ are constants, $\sum_j (AS)_{jk} = \sum_k (AS)_{jk} = 0$

$(PAS)_{ijk}$ are constants, $\sum_i (PAS)_{ijk} = \sum_j (PAS)_{ijk} = \sum_k (PAS)_{ijk} = 0$

If S = RANDOM

$$s_k = N(o, \sigma_S^2); \quad (Ps)_{ik} = N(o, \sigma_{Ps}^2)$$

$$(As)_{jk} = N(o, \sigma_{As}^2); \quad (PAs)_{ijk} = N(o, \sigma_{PAs}^2)$$

s_k, $(Ps)_{ik}$, $(AS)_{jk}$, $(PAs)_{ijk}$ are uncorrelated.

Define effect γ and its interactions:

$$\gamma_k = s_k + (Ps)_{.k} + (As)_{.k} + (PAs)_{..k}$$

$$(P\gamma)_{ik} = \left[(Ps)_{ik} - (Ps)_{.k}\right] + \left[(PAs)_{i\cdot k} - (PAs)_{..k}\right],$$

$$\sum_i (P\gamma)_{ik} = 0$$

$$(A\gamma)_{jk} = \left[(As)_{jk} - (As)_{.k}\right] + \left[(PAs)_{.jk} - (PAs)_{..k}\right],$$

$$\sum_j (A\gamma)_{jk} = 0$$

$$(PA\gamma)_{ijk} = (PAs)_{ijk} - (PAs)_{i\cdot k} - (PAs)_{.jk} + (PAs)_{..k}$$

$$\sum_i (PA\gamma)_{ijk} = \sum_j (PA\gamma)_{ijk} = 0$$

$$\text{Time on Target} = M + P_i + A_j + \gamma_k + (PA)_{ij} + (P\gamma)_{ik} + (A\gamma)_{jk} + (PA\gamma)_{ijk} + r_{ijkl}$$

The corresponding Analysis of Variance table for the above model with S = random is given by Exhibit 1. Exhibit 2 indicates the contrast in the numerical results for S = fixed and the results for S = random.

EXHIBIT 1

ANOVA SUBJECTS RANDOM (RUNS RANDOM)

SV	DF	EXPECTED MEAN SQUARES		ET
P	1	$180(\frac{2}{I})\,\sigma_P^2 + 30(\frac{2}{I})\,\sigma_{P\gamma}^2$	$+\sigma_r^2$	$P\gamma$
A	2	$120(\frac{3}{J})\,\sigma_A^2 + 20(\frac{3}{J})\,\sigma_{A\gamma}^2$	$+\sigma_r^2$	$A\gamma$
γ	5	$60\,\sigma_\gamma^2$	$+\sigma_r^2$	r
PA	2	$60(\frac{2}{I})(\frac{3}{J})\,\sigma_{PA}^2 + 10(\frac{2}{I})(\frac{3}{J})\,\sigma_{PA\gamma}^2$	$+\sigma_r^2$	$PA\gamma$
$P\gamma$	5	$30(\frac{2}{I})\sigma_{P\gamma}^2$	$+\sigma_r^2$	r

EXHIBIT 1 (cont.)

SV	DF	EXPECTED MEAN SQUARES		ET
$A\gamma$	10	$20(\frac{3}{2})\sigma^2_{A\gamma}$	$+\sigma^2_r$	r
$PA\gamma$	10	$10(\frac{2}{1})(\frac{3}{2})\sigma^2_{PA\gamma}$	$+\sigma^2_r$	r
r/PAs	324		$+\sigma^2_r$	
	359			

EXHIBIT 2

ANALYSIS OF VARIANCE

SV	DF	SS	MS	F(S Fixed)	F(S Random)
P	1	1,876.90	1,876.90	57.01**	3.41
A	2	1,148.69	574.35	17.45**	13.01**
S,γ	5	7,938.99	1,587.80	48.23**	48.23**
PA	2	225.87	112.94	3.43*	2.86
PS,Pγ	5	2,753.63	550.73	16.73**	16.73**
AS,Aγ	10	441.41	44.14	1.34	1.34
PAS,PAγ	10	395.10	39.51	1.20	1.20
r/PAS	324	10,665.40	32.92	--	--
Total	359	25,445.99			

Using the Newman-Keuls Multiple Comparisons Test and Satterthwaite's approximation for chi-square, one obtains the statistical results of the experimental data given in Exhibits 3 through 13.

EXHIBIT 3

95% CONFIDENCE LIMITS FOR σ^2_r

$$\hat{\sigma}^2_r = 32.92$$

$$\frac{\hat{\sigma}^2_r}{(\frac{\chi^2}{df}).975} < \sigma^2_r < \frac{\hat{\sigma}^2_r}{(\frac{\chi^2}{df}).025}$$

EXHIBIT 3 (cont.)

$$\left(\frac{\chi^2}{324}\right)_{.975} = 1.10 \qquad \left(\frac{\chi^2}{324}\right)_{.025} = 0.912$$

$$29.93 < \sigma_r^2 < 36.10$$

$$5.\overline{5} < \sigma_r < 6.0$$

EXHIBIT 4

MAIN EFFECTS OF P

Protected = 20.79*
Unprotected = 16.22

$$\begin{pmatrix} P_1 = 2.28 \\ P_2 = -2.28 \end{pmatrix}$$

$P_2 < P_1$ (S FIXED)

$$\Pr\ (3.36 < P_1 - P_2 < 5.76) = .95$$

$\underline{P_2,\ P_1}$ (S RANDOM)

$$\Pr\ (-0.30 < P_1 - P_2 < 9.44) = .95$$

*Grand Mean = 18.51

EXHIBIT 5

MAIN EFFECTS OF A

Control 1 = 20.82
3G = 16.47
Control 2 = 18.23

$$\begin{pmatrix} A_1 = 2.31 \\ A_2 = -2.04 \\ A_3 = -0.28 \end{pmatrix}$$

3G	C2	C1
16	18	21

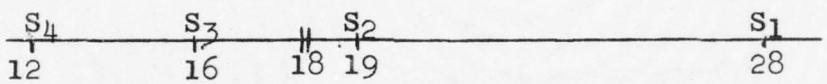
$G < C_2 < C_1$ (S FIXED)
$\underline{G,\ C_2} < C_1$ (S RANDOM)
} Newman-Keuls Test

Grand Mean = 18.51

EXHIBIT 6

MAIN EFFECTS OF S (S FIXED)

		(S)	
$S_1 = 27.87$	$S_4 = 12.48$	9.36	-6.03
$S_2 = 18.73$	$S_5 = 17.88$	0.22	-0.63
$S_3 = 15.68$	$S_6 = 18.38$	-2.83	-0.13



$$S_4 < S_3 < \underline{S_5, S_6, S_2} < S_1$$

Grand Mean = 18.51

EXHIBIT 7

APPROX.* 95% CONFIDENCE LIMITS FOR σ_γ^2 (S RANDOM)

$$\hat{\sigma}_\gamma^2 = \frac{MS(\gamma) - MS(r)}{60} = \frac{1587.80 - 32.92}{60} = 25.91$$

$$\frac{\hat{\sigma}_\gamma^2}{\left(\frac{\chi^2}{df}\right)_{.975}} < \sigma_\gamma^2 < \frac{\hat{\sigma}_\gamma^2}{\left(\frac{\chi^2}{df}\right)_{.025}}$$

$$\left(\frac{\chi^2}{5}\right)_{.975} = 2.57$$

$$\left(\frac{\chi^2}{5}\right)_{.025} = 0.166$$

$$10 < \sigma_\gamma^2 < 156$$

$$3.3 < \sigma_\gamma < 12.5$$

* Using Satterthwaite's Approximate chi-square.

EXHIBIT 8

COMPARISON OF INDIVIDUAL PERFORMANCE

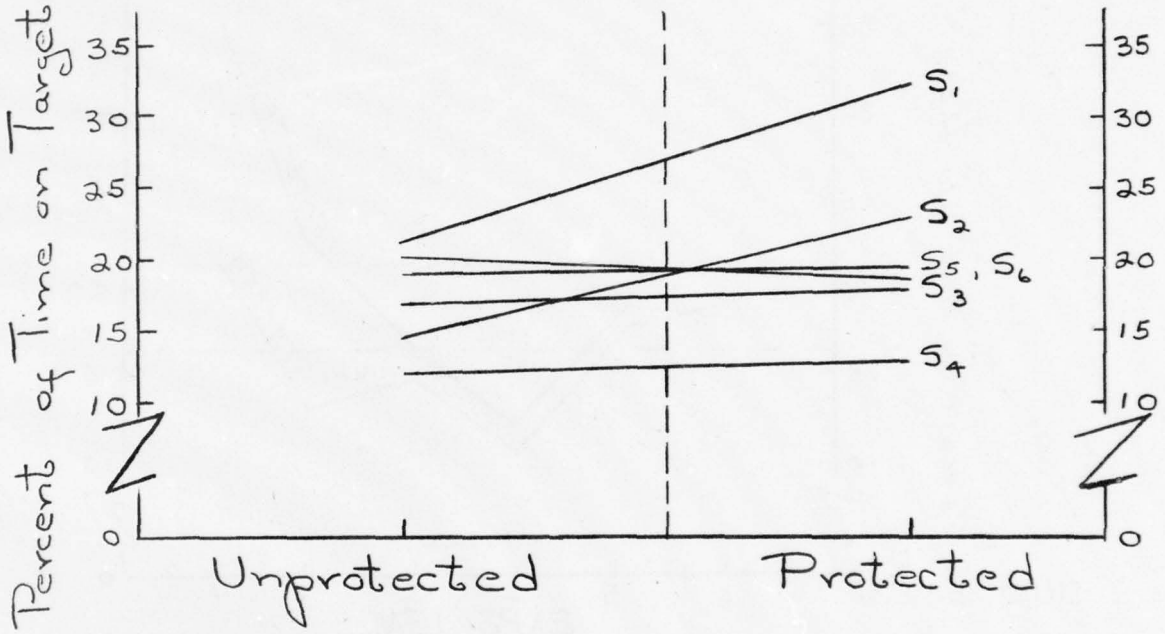
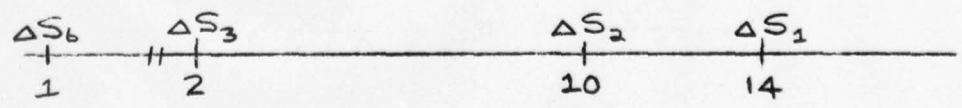


EXHIBIT 9

PS INTERACTION (S FIXED)

$$\Delta S_1 = (P_1-P_2)\ S_1 = 14.27$$
$$\Delta S_2 = (P_1-P_2)\ S_2 = 9.93$$
$$\Delta S_3 = (P_1-P_2)\ S_3 = 1.77$$
$$\Delta S_4 = (P_1-P_2)\ S_4 = 1.37$$
$$\Delta S_5 = (P_1-P_2)\ S_5 = 0.90$$
$$\Delta S_6 = (P_1-P_2)\ S_6 = -0.83$$

$$\begin{pmatrix} (PS) & \\ 4.85 & -4.85 \\ 2.68 & -2.68 \\ -1.40 & 1.40 \\ -1.60 & 1.60 \\ -1.83 & 1.83 \\ -2.70 & 2.70 \end{pmatrix}$$



$\Delta S_6, \Delta S_5, \Delta S_4, \Delta S_3 < \Delta S_2, \Delta S_1$

EXHIBIT 10

EFFECT OF PRESSURE PROTECTION

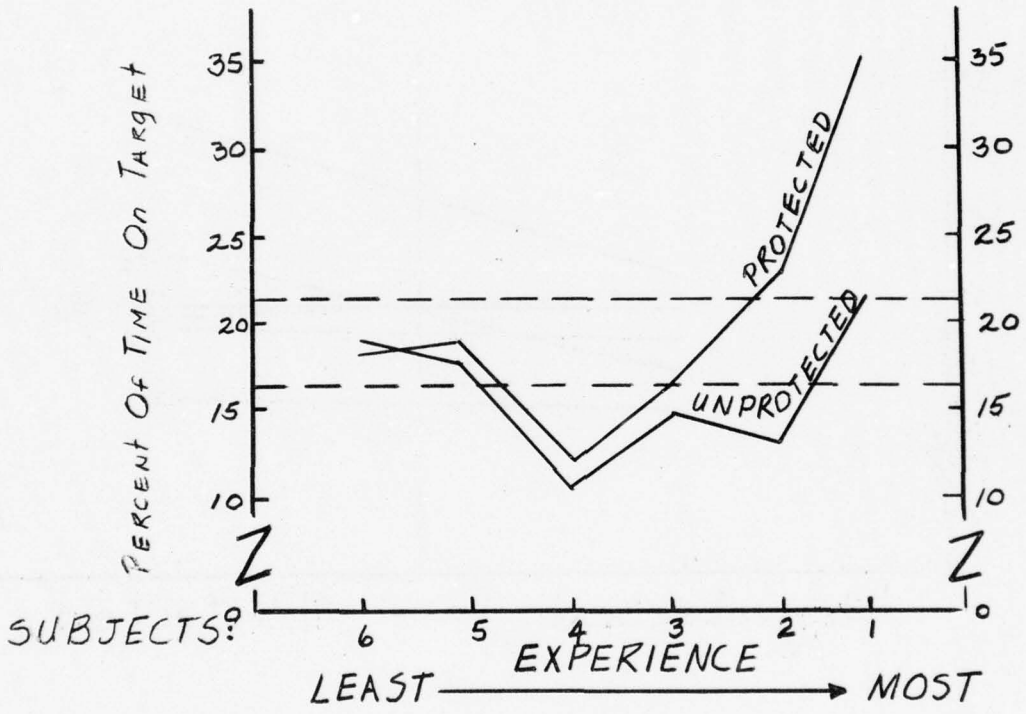


EXHIBIT 11

APPROXIMATE 95 PERCENT CONFIDENCE LIMITS

FOR $\sigma^2_{P\gamma}$ (S RANDOM)

$$\hat{\sigma}^2_{P\gamma} = \frac{1}{2}\,\frac{MS(P\gamma) - MS(r)}{30} = \frac{1}{2}\,\frac{550.73 - 32.92}{30} = 8.63$$

$$\frac{\hat{\sigma}^2_{P\gamma}}{2.57} < \sigma^2_{P\gamma} < \frac{\hat{\sigma}^2_{P\gamma}}{0.166}$$

$$3 < \sigma^2_{P\gamma} < 52$$

$$1.7 < \sigma_{P\gamma} < 7.2$$

EXHIBIT 12

CONTROL VS ACCELERATION

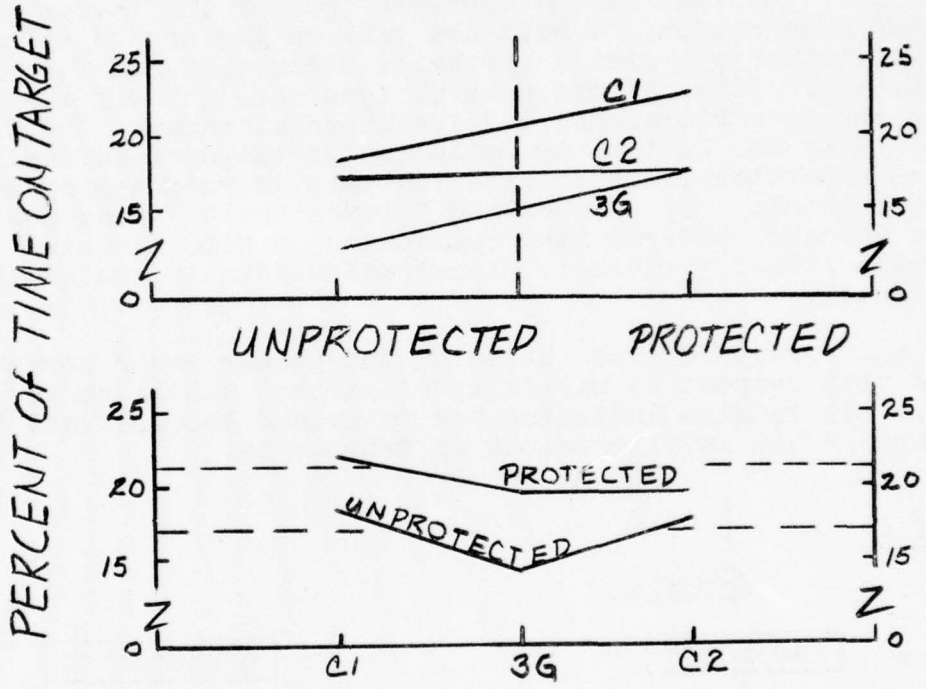


EXHIBIT 13

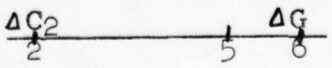
PA INTERACTION

$\Delta C_1 = (P_1-P_2)C_1 = 5.30^*$

$\Delta G = (P_1-P_2)G = 6.03$

$\Delta C_2 = (P_1-P_2)C_2 = 2.37$

$$\begin{matrix} & (PA) \\ \begin{pmatrix} 0.37 & -0.37 \\ 0.73 & -0.73 \\ -1.10 & 1.10 \end{pmatrix} \end{matrix}$$



$\Delta C_2 < \underline{\Delta C_1, \Delta G}$ (S FIXED)

$\underline{\Delta C_2, \Delta C_1, \Delta G}$ (S RANDOM)

*$P_1-P_2=4.56$

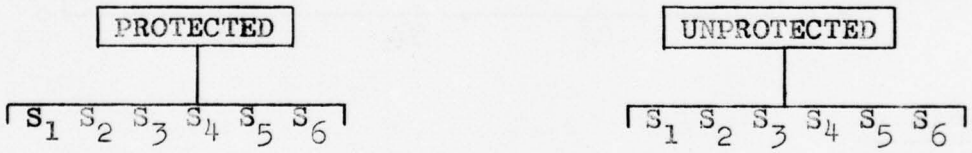
IV. PARTIALLY HIERARCHAL MODELS

The example I have just finished talking about is a factorial model. It is called "mixed" whenever some of the factors are fixed and some random. I will now take up the second main point, which is to describe what a partially hierarchal model is and what it does. It is a more general type and includes the factorial and the pure hierarchal model as special cases. The term was coined by Dr. Harter, my colleague at the Aeronautical Research Laboratory, and I think the term is very appropriate for these models. Dr. Harter and I investigated these models and our research efforts have culminated in WADC Technical Report 55-33 titled "Partially Hierarchal Models in the Analysis of Variance".

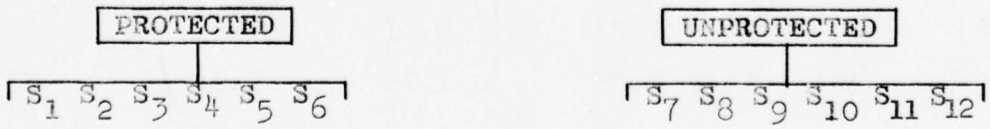
In our report we give tables of the proper error terms for F-tests with respect to partially hierarchal models up to four factors. It is also indicated how to extend the analysis to n factors. This is illustrated in Exhibit 14.

EXHIBIT 14

S FACTORIAL



S NESTING WITHIN P



(G_1) (G_2)

Suppose now that 2 groups of 6 subjects are used on the experiment. The first group of 6 subjects, G_1, performs the experiment with protection and the second group of 6 subjects, G_2, performs the experiment without protection; i.e. a given subject performs only <u>one</u> level of the protection factor.

Thus each subject no longer occurs with all levels of P as in a factorial model. In such a situation, the subjects factor is a nesting factor, and it is said to "nest within the P factor". This can also be described as "subjects within protection". The main characteristic of the nesting is that there is no longer a direct relationship between the first (ith) subject with protection and the first (ith) subject without protection.

Definition: A pure hierarchal model is one where all the factors form a nested set; whereas a partially hierarchal model is one in which some factors are of the nesting type and the remaining factors are factorial.

Thus, the centrifuge experiment as now altered is a partially hierarchal experiment (instead of a factorial) where the Subjects factor is nested within the Protection factor. A mathematical model for this partially hierarchal case is given in Exhibit 15A and 15B with the corresponding Analysis of Variance table given in Exhibit 16. Exhibit 17 shows a comparison of the values of F for the factorial model (taken from Exhibit 2) and the values of F for the partially hierarchal model (using the same data as if it originated from the partially hierarchal situation).

EXHIBIT 15A

MATHEMATICAL MODEL FOR S(RANDOM) NESTING IN P

$$\text{Time on Target} = M + P_i + A_j + s^{*}_{ik} + (PA)_{ij} + (As)^{*}_{ijk} + r_{ijkl}$$

M = constant $\qquad P_1 = -P_2 = \text{constant}$

c_1, 3G, c_2 are constants, $\sum_j A_j = 0$

$(PA)_{ij}$ are constants, $\sum_i (PA)_{ij} = 0 = \sum_j (PA)_{ij}$

$s_{ik} = N(0, \sigma^2_s)$; $(As)_{ijk} = N(0, \sigma^2_{As})$; $r_{ijkl} = N(0, \sigma^2_r)$

s_{ik}, $(As)_{ijk}$, r_{ijkl} are mutually uncorrelated

$^{*} s_{ik} = s_k + (Ps)_{ik}$; $(As)_{ijk} = (As)_{jk} + (APs)_{ijk}$

EXHIBIT 15B

DEFINITION OF EFFECT γ AND CORRESPONDING INTERACTIONS

$$\gamma_{ik} = s_{ik} + (As)_{i\cdot k}$$

$$(A\gamma)_{ijk} = (As)_{ijk} - (As)_{i\cdot k}, \quad \sum_j (A\gamma)_{ijk} = 0$$

$$\text{Time on Target} = M + P_i + A_j + \gamma_{ik} + (PA)_{ij} + (A\gamma)_{ijk} + r_{ijkl}$$

EXHIBIT 16

ANOVA PARTIALLY HIERARCHAL S(RANDOM) NESTING IN P

SV	DF	EXPECTED MEAN SQUARE			ET
P	1	$180(\frac{2}{1})\,\sigma^2_P + 30\,\sigma^2_\gamma$		$+\sigma^2_r$	γ
A	2	$120(\frac{3}{2})\,\sigma^2_A$	$+ 10(\frac{3}{2})\,\sigma^2_{A\gamma}$	$+\sigma^2_r$	$A\gamma$
$r\|P$	10	$30\,\sigma^2_\gamma$		$+\sigma^2_r$	r
PA	2	$60(\frac{3}{2})(\frac{2}{1})\sigma^2_{PA}$	$+10(\frac{3}{2})\sigma^2_{A\gamma}$	$+\sigma^2_r$	$A\gamma$
$A\gamma\|P$	20		$10(\frac{3}{2})\sigma^2_{A\gamma}$	$+\sigma^2_r$	r
$r\|PAs$	324			σ^2_r	---
Total	359				

EXHIBIT 17

ANOVA S NESTING IN P

SV	DF	SS	MS	F^1	F^2
P	1	1877	1877	1.75	3.41
A	2	1149	574	13.73**	13.01**
s/P	10	10,693	1069	32.48**	
s	5	7939	1588		48.23**
P s	5	2754	551		16.73**
PA	2	226	113	2.70	2.86
As/P	20	837	41.8	1.27	
As	10	441	44.1		1.34
PAs	10	395	39.5		1.20
r/PAs	324	10,665	32.9	--	--
Total	359	25,446			

1 Partially Hierarchal

2 Factorial

Possible advantages of using a partially hierarchal model (and an experiment based on it) are:

(1) it is often easier to obtain a large number of subjects for short periods of time;

(2) learning or fatigue effects, if they are important, are avoided.

Hidden Pitfalls in such a model are:

(Error 1)	P_1	P_2	
If:	15	1	$(P_1 >> P_2)$
	$\frac{1}{16}$	$\frac{15}{16}$	$(G_1 << G_2)$

Conclusion: Get no sig. P effect when actually $P_1 >> P_2$

(Subject group 1 << Subject group 2)

(Error 2) P_1 P_2 Conclusion: Get $P_1 << P_2$ sig.,

If: 1 1 ($P_1 = P_2$) when actually $P_2 \equiv P_1$

$\frac{1}{2}$ $\frac{15}{16}$ ($G_1 << G_2$) ($G_1 << G_2$)

(Error 3) P_1 P_2

If: 5 1 $P_1 > P_2$ Conclusion: get $P_1 < P_2$

$\frac{1}{6}$ $\frac{15}{16}$ $G << G_2$ when actually $P_1 > P_2$

($G_1 << G_2$)

(Error 4) Even if G_1 and G_2 differ in same direction as P_1 and P_2, conclusion may not be quite right.

If: P_1 P_2

i.e., 15 1 $P_1 >> P_2$ Conclusion: $G_1 >> G_2$ inflates (or deflates as the case may be) the difference between P_1 and P_2.

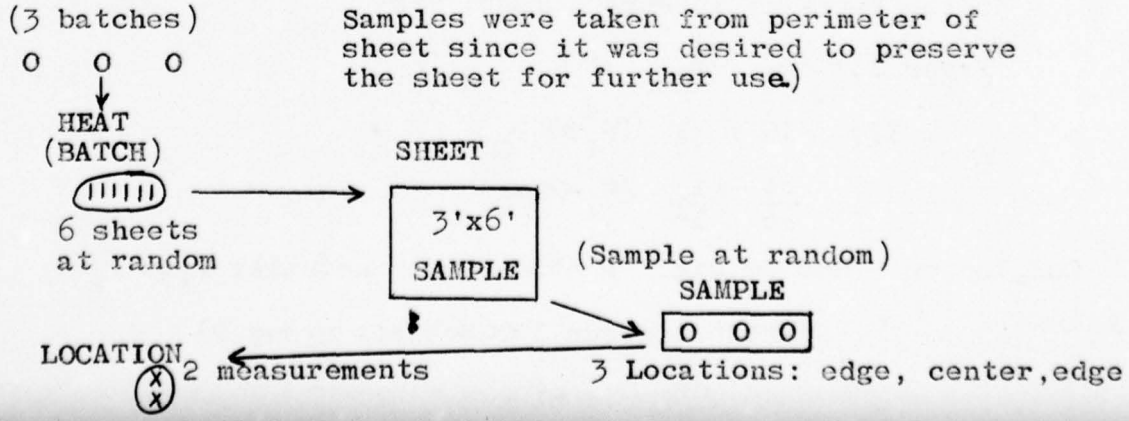
$\frac{20}{35}$ $\frac{1}{2}$ $G_1 >> G_2$

One method to insure against these errors if possible is to pre-test the subjects for assurance that the two groups are essentially alike.

Another example of a partially hierarchal model is an experiment being conducted by the Metals Branch, Materials Laboratory, at Wright Air Development Center. The experiment was to be performed to evaluate the mill production quality with respect to uniformity of roll sheets of metal.

The purpose was to make inferences concerning the uniformity of rolls with respect so some measured characteristic in the metal.

The following diagram indicates the manner in which the measurements were made.



Total number of measurements = 3 X 6 X 3 X 2 = 108

A mathematical model for this partially heierarchal experiment is given in Exhibit 18 and 19 with the corresponding Analysis of Variance table given in Exhibit 20.

EXHIBIT 18

METALS EXPERIMENT

H = Heat(Batch) S = Sheet - Sample

L = Location M = Measurement

H	----	(random)
S|H	nesting in H	(random)
L	factorial	(fixed)
M|LSH	nesting in L, S	(random)

EXHIBIT 19A

MATHEMATICAL MODEL FOR PARTIALLY HIERARCHAL

$$X_{ijkr} = M + h_i + s^{*}_{ij} + L_k + (hL)_{ik} + (sL)^{*}_{ijk} + m_{ijkr}$$

L_1, L_2, L_3 = constants, $\sum_k L_k = 0$

$h_i = N(0, \sigma^2_h)$; $s_{ij} = N(0, \sigma^2_s)$; $(hL)_{ik} = N(0, \sigma^2_{hL})$

$(sL)_{ijk} = N(0, \sigma^2_{sL})$; $m_{ijkr} = N(0, \sigma^2_m)$

h_i, s_{ij}, $(hL)_{ik}$, $(sL)_{ijk}$, m_{ijkr} are mutually uncorrelated.

*$s_{ij} = s_j + (hs)_{ij}$; $(sL)_{ijk} = (sL)_{jk} + (hsL)_{ijk}$

EXHIBIT 19B

DEFINITION OF EFFECTS a, β

AND CORRESPONDING INTERACTIONS

$$a_i = h_i + (hL)_{i\cdot} \quad , \quad \beta_{ij} = s_{ij} + (sL)_{ij\cdot}$$

$$(aL)_{ik} = (hL)_{ik} - (hL)_{i\cdot} \quad , \quad \sum_k (aL)_{ik} = 0$$

$$(\beta L)_{ijk} = (sL)_{ijk} - (sL)_{ij\cdot} \quad , \quad \sum_k (\beta L)_{ijk} = 0$$

$$X_{ijkr} = M + a_i + \beta_{ij} + L_k + (aL)_{ik} + (\beta L)_{ijk} + {}^{m}_{ijkr}$$

EXHIBIT 20

ANOVA PARTIALLY HIERARCHAL MODEL

SV	DF	EXPECTED MEAN SQUARE	ET
a	2	$36\,\sigma^2_a + 6\,\sigma^2_\beta + \sigma^2_m$	β
$\beta \mid a$	15	$6\,\sigma^2_\beta + \sigma^2_m$	m
L	2	$36(\frac{3}{2})\,\sigma^2_L + 12(\frac{3}{2})\,\sigma^2_{aL} + 2(\frac{3}{2})\,\sigma^2_{\beta L} + \sigma^2_m$	aL
aL	4	$12(\frac{3}{2})\,\sigma^2_{aL} + 2(\frac{3}{2})\,\sigma^2_{\beta L} + \sigma^2_m$	βL
$\beta L \mid a$	30	$2(\frac{3}{2})\,\sigma^2_{\beta L} + \sigma^2_m$	m
m | Lsh	54	σ^2_m	--
Total	107		

Questions that can be answered by use of this partially hierarchal model:

1. Do batches differ significantly as compared with variations in sheets within a given batch?

2. For a given batch do sheets differ significantly as compared with variations in measurements?

3. Do locations show a significant difference as compared to the variation in location differences from batch to batch?

4. Do the variations in location differences from batch to batch differ significantly as compared to the variations in location differences from sheet to sheet within a batch?

5. Are variations in location differences from sheet to sheet within a batch significantly different or can they be attributed to the errors in measurement?

REFERENCES

(1) Keuls, M., "The Use of the 'Studentized Range' in Connection With an Analysis of Variance", Euphytica 1 (1952): 112-122.

(2) Satterthwaite, F. E., "An Approximate Distribution of Estimates of Variance Components", Biometrics 2 (1946): 110-114.

TWO-SIDED TOLERANCE LIMITS FOR NORMAL DISTRIBUTIONS USING THE RANGE*

By George J. Resnikoff

INTRODUCTION

The quality of manufactured product is often specified by giving a range or interval for a measureable characteristic. The upper and lower limits to this interval are called tolerance limits. These limits are such that the probability is equal to a preassigned value that the interval includes at least a specified proportion of the statistical universe.

The problem of computing two-sided tolerance limits on the basis of a sample is as follows:

Let x be a random variable, with distribution function F, and let $x_1, x_2, \ldots, x_N$ be a sample of N observations on x. It is required to construct two functions of the sample, L_1 and L_2, such that the probability is γ that at least a specified proportion P of the distribution is included between L_1 and L_2. The limits L_1 and L_2 are called tolerance limits. γ is called the confidence coefficient.

TWO-SIDED TOLERANCE LIMITS FOR NORMAL DISTRIBUTIONS

For the case of a normal distribution with mean μ and standard deviation σ, both parameters unknown, Wald and Wolfowitz have given an excellent approximation to the problem of setting two-sided tolerance limits [14].

Let $x_1, x_2, \ldots, x_N$ be a sample of N observations from the normal distribution. Let

$$\bar{x} = \frac{\sum x_i}{N} \quad \text{and} \quad s^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}$$

Define

$$A(\bar{x}, s, \lambda) = \frac{1}{\sqrt{2\pi}} \int_{\bar{x} - \lambda s}^{\bar{x} + \lambda s} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt.$$

*This work was supported in part by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

$A(\bar{x}, s, \lambda)$ is the proportion of the normal distribution included between the limits $\bar{x} + \lambda s$ and $\bar{x} - \lambda s$.

Wald and Wolfowitz have shown that a very good approximation to the value of λ such that $\Pr\{A(\bar{x}, s, \lambda) > P\} = \gamma$ is given by

$$\lambda^* = r\sqrt{\frac{N-1}{x^2_\gamma}}$$

where x^2_γ is that value of a chi-square variable with N-1 degrees of freedom which is exceeded with probability γ, and r is the solution of the equation

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{1}{\sqrt{N}} - r}^{\frac{1}{\sqrt{N}} + r} e^{-\frac{t^2}{2}} \, dt = P$$

Utilizing this approximation A. H. Bowker computed extensive tables of factors K such that $\bar{x}$-Ks and $\bar{x}$+Ks are two-sided tolerance limits. These tables are given in [1] . Bowker also has given an asymptotic solution to this case of setting two-sided tolerance limits for a normal universe [2].

TWO-SIDED TOLERANCE LIMITS FOR THE NORMAL DISTRIBUTION USING THE SAMPLE RANGE OR SAMPLE AVERAGE RANGE

Let $\hat{\sigma}$ be a sample estimate of σ which is independent of the sample mean x, and such that

$$A(\bar{x}, \hat{\sigma}, \lambda) = \frac{1}{\sqrt{2\pi}} \int_{\bar{x} - \lambda\hat{\sigma}}^{\bar{x} + \lambda\hat{\sigma}} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} \, dt$$

is a strictly increasing function of $\hat{\sigma}$. Then it follows directly from the arguments of Wald and Wolfowitz cited in Section 2 above, that an approximation to the value of λ such that $\Pr\{A(\bar{x}, \hat{\sigma}, \lambda) > P\} = \gamma$ is given by

$$\lambda^* = r/\hat{\sigma}_\gamma$$

where $\hat{\sigma}_\gamma$ is such that $\Pr\{\hat{\sigma} < \hat{\sigma}_\gamma\} = 1 - \gamma$, and r is, as before, the root of the equation

$$\frac{1}{\sqrt{2\pi}} \int_{-r}^{+r} e^{-\frac{t^2}{2}} dt = P.$$

The accuracy of the approximation is the same regardless of whether the sample standard deviation s is used or whether some less efficient statistic $\hat{\sigma}$ is used in place of s. The effect of using $\hat{\sigma}$ instead of s is that, on the average, somewhat wider tolerance intervals will result.

Among such sample estimates of σ are the sample range and the sample average-range. Let $x^{(N)}$ denote the largest observation in the sample and $x^{(1)}$ the smallest, then the sample range $R = x^{(N)} - x^{(1)}$. For N = 2, R and s differ only by a multiplicative constant. For N larger than 10 the efficiency of R as compared with s decreases rapidly. It is customary for large samples to divide the sample randomly into m equal groups of size n and to compute the range of each group. The average of these m ranges is called the sample average-range. We shall denote this statistic by $\bar{R}_{m,n}$. The total sample size N = mn. For m = 1, $R_{1,n}$ is the sample range, R, for a single sample of size $N = n$. For conciseness in the subsequent discussion we shall refer to both the sample range and the sample average-range as $\bar{R}_{m,n}$.

The probability integral of $\bar{R}_{1,n}$ from a normal population with variance σ^2, for $n = 2, 3, \ldots, 20$, has been tabled by Pearson and Hartley [6]. For m = 2, 3, ..., 10, and for n = 5, the probability distribution and percentage points of $\bar{R}_{m,5}/\sigma$ have been computed by the writer and are given in [7]. For sample sizes which are multiples of 5, $\bar{R}_{m,5}$ is used extensively in industrial applications.

THE TABLES OF FACTORS FOR TOLERANCE LIMITS BASED ON THE RANGE OR AVERAGE-RANGE

Factors K such that $\bar{x} - K\bar{R}_{1,n}$ and $\bar{x} + K\bar{R}_{1,n}$ are two-sided tolerance limits for a normal universe are given in Table 1 for $n = 2, 3, \ldots, 10$; for $\gamma = .75, .90,$

.95, .975, .99, and .995. Values of P are .75, .90, .95, .975, .99, .995, and .999.

Factors K such that $\bar{x} - K\bar{R}_{m,5}$ and $\bar{x} + K\bar{R}_{m,5}$ are two-sided tolerance limits for a normal universe are given in Table 2 for m = 2, 3, ..., 10; for γ = .75, .90, .95, .975, .99, and .995. Values of P are .75, .90, .95, .975, .99, .995, and .999.

The factors K used in Table 1 were obtained by solving for K in

$$K = \frac{r}{R_\gamma}$$

where R_γ is such that $\Pr\{\bar{R}_{1,n} > R_\gamma\} = \gamma$. The values of R_γ were obtained by inverse interpolation in the Pearson and Hartley tables [6] for n = 3, 4, ..., 10. For n = 2 the values of R_γ were obtained from tables of percentage points of the X^2 distribution. For samples of size 2 the range is the same as the sample standard deviation, except for the factor $\sqrt{2}$.

The factors K used in Table 2 were obtained by solving for K in

$$K = \frac{r}{R'_\gamma}$$

where R'_γ is such that $\Pr\{\bar{R}_{m,5} > R'_\gamma\} = \gamma$. The values of R'_γ were obtained from the tables of the percentage points of the distribution of the average-range for subgroups of size 5, computed earlier by the writer [7].

The values of r which are solutions to the equation

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{1}{\sqrt{N}} - r}^{\frac{1}{\sqrt{N}} + r} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} = P$$

were obtained by the use of Newton's Method on the equation defining r.

Examples of the use of these tables follow in the two subsequent paragraphs.

In the manufacture of electron tubes to be used in stable amplifiers it is desired to know, with confidence coefficient of .99, limits within which 90 percent of the future tube transconductances lie. The required tests are made on 8 tubes and the transconductances, observed in micromhos, are as follows:

4430

4287

4450

4295

4340

4407

4295

4388

4356

From Table 1, the value of K corresponding to n = 9, γ = .99 and P = .90 is 1.290. For this example, $\bar{x}$ is found to be 4349.8 and $\bar{R}_{1,9}$ is 163. Thus the tolerance limits are given by $4349.8 \pm (1.290)(163) = [4139.53, 4560.07]$.

In studying another characteristic of the vacuum tubes, a sample of size N = 20 was taken. The sample was grouped into 4 subsamples of 5 observations each. The ranges of each of the 4 subgroups were found to be

36.02

37.45

36.95

36.30

The average $\bar{x}$ of the 20 observations was given as 448.50 and the average range $\bar{R}_{4,5}$ is computed from the above and found to be 36.68.

Assuming $\gamma = .995$ and $P = .90$, from Table 2 we obtain the value of $K = 1.283$ corresponding to $m = 4$. Thus the tolerance limits are given by $448.50 \pm (1.283)(36.68) = [401.44, 495.56]$.

After the present tables were computed, there appeared in the March 1957 issue of the Journal of the American Statistical Association a set of tables by S. K. Mitra, under the title, "Tables for Tolerance Limits for a Normal Population Based on Sample Mean and Range or Mean Range".

The tables included herein duplicate some of the results given by Mitra. However, it was decided to proceed with the publication of the present paper for the following reasons:

1) The present tables use the exact distribution of the mean-range in their construction whereas Mitra's tables compound two approximations. In particular the expansion used by Mitra for computing the percentage points of a χ^2 variate is poor for small values of degrees of freedom, especially in the tails of the distribution. As a result the tolerance factors based on an average range statistic from m subgroups of 5 observations each, given in the present report, will give more correct results than will Mitra's tables, for $m = 4, 5, \ldots, 10$.

2) The present tables include tolerance factors based on the average-range statistic of m groups of 5 observations each, for the useful cases of $m = 2$ and 3, which are not included in Mitra's tables.

3) The present tables include some values of confidence coefficient and population proportion not included in Mitra's tables.

REFERENCES

1. Bowker, Albert H.: "Computation of Factors for Tolerance Limits on a Normal Distribution When the Sample Is Large," Annals of Mathematical Statistics, Vol. 17 (1946), pp. 238-240.

2. Bowker, Albert H.: "Tolerance Limits for Normal Distributions," Techniques of Statistical Analysis, edited by Eisenhart, Hastay and Wallis, McGraw-Hill, New York, 1947, Chapter 2.

3. Johnson, N. L., and Welch, B. L.: "Applications of the Non-central t-Distribution," Biometrika, Vol. 31 (1940), pp. 362-389.

4. Patnaik, P. B.: "The Use of Mean Range in Statistical Tests," Biometrika, Vol. 37 (1950) p. 78

5. Paulson, Edward: "A Note on Tolerance Limits", Annals of Mathematical Statistics, Vol. 14 (1943), pp. 90-93.

6. Pearson, E. S., and Hartley, H. O.: "The Probability Integral of the Range in Samples of n observations from a Normal Population," Biometrika, Vol. 32 (1942) pp. 301-310.

7. Resnikoff, George J.: "The Distribution of the Average-Range for Subgroups of Five," Technical Report No. 15, Contract N60nr-25126, Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California, 1954.

8. Resnikoff, George J. and Lieberman, Gerald J.: "Tables of the Non-Central t-Distribution," Stanford University Press, Stanford, California (1957) 736 pp.

9. Robbins, Herbert: "On Distribution-free Tolerance Limits in Random Sampling," Annals of Mathematical Statistics, Vol. 15 (1944), pp. 214-216.

10. Scheffe, H., and Tukey, J. W.: "A Formula for Sample Sizes for Population Tolerance Limits," Annals of Mathematical Statistics, Vol. 15 (1944), p. 217.

11. Shewhart, Walter A., with the editorial assistance of Deming, W. Edwards: "Statistical Method from the Viewpoint of Quality Control," Graduate School, U. S. Department of Agriculture, Washington, D. C., 1939 155 + ix pp.

12. Wald, Abraham: "Setting of Tolerance Limits When Sample is Large", Annals of Mathematical Statistics, Vol. 13 (1942), pp. 389-399.

13. Wald, Abraham: "An Extension of Wilks' Method of Setting Tolerance Limits," Annals of Mathematical Statistics, Vol. 14 (1943), pp. 45-55.

14. Wald, Abraham and Wolfowitz, J.: "Tolerance Limits for a Normal Distribution," Annals of Mathematical Statistics, Vol. 17 (1946), pp. 208-215.

15. Wilks, S. S.: "Determination of Sample Sizes for Setting Tolerance Limits," Annals of Mathematical Statistics, Vol. 12 (1941), pp. 91-96.

16. Wilks, S. S.: "Statistical Prediction with Special Reference to the Problem of Tolerance Limits," Annals of Mathematical Statistics, Vol. 13 (1942), pp. 400-409.

Table 1. Tolerance Factors for Normal Distributions Utilizing the Range

Factors K for two-sided tolerance limits such that the probability is γ that at least P of the distribution will be included between $\bar{x} \pm K\bar{R}_{1,n}$, where $\bar{R}_{1,n}$, is the sample range for a sample of size n.

	γ = .75						
n \ P	.75	.90	.95	.975	.99	.995	.999
2	3.181	4.456	5.243	5.932	6.739	7.290	8.429
3	1.312	1.857	2.197	2.498	2.850	3.092	3.592
4	0.916	1.301	1.544	1.759	2.011	2.185	2.545
5	0.744	1.060	1.259	1.436	1.645	1.788	2.087
6	0.647	0.923	1.097	1.252	1.435	1.561	1.823
7	0.584	0.833	0.992	1.132	1.299	1.413	1.652
8	0.540	0.771	0.917	1.048	1.202	1.308	1.530
9	0.507	0.723	0.861	0.984	1.129	1.229	1.438
10	0.481	0.687	0.817	0.934	1.072	1.168	1.366

	γ = .90						
n \ P	.75	.90	.95	.975	.99	.995	.999
2	8.066	11.298	13.294	15.043	17.089	18.486	21.373
3	2.170	3.071	3.634	4.130	4.713	5.113	5.940
4	1.322	1.878	2.228	2.538	2.903	3.154	3.674
5	1.003	1.428	1.696	1.935	2.216	2.409	2.811
6	0.838	1.194	1.420	1.620	1.858	2.021	2.360
7	0.735	1.049	1.248	1.425	1.635	1.779	2.080
8	0.666	0.951	1.132	1.292	1.483	1.614	1.888
9	0.615	0.879	1.046	1.195	1.372	1.494	1.747
10	0.577	0.824	0.981	1.121	1.287	1.401	1.640

Table 1. Tolerance Factors for Normal Distributions Utilizing the Range (continued)

	$\gamma = .95$						
n \ p	.75	.90	.95	.975	.99	.995	.999
2	16.163	22.641	26.640	30.145	34.245	37.045	42.831
3	3.112	4.403	5.210	5.922	6.758	7.331	8.517
4	1.705	2.423	2.874	3.274	3.745	4.068	4.738
5	1.229	1.749	2.078	2.370	2.715	2.952	3.444
6	0.995	1.418	1.686	1.924	2.206	2.400	2.803
7	0.856	1.221	1.453	1.659	1.903	2.071	2.420
8	0.763	1.090	1.297	1.481	1.700	1.850	2.164
9	0.698	0.997	1.186	1.355	1.555	1.694	1.981
10	0.648	0.926	1.103	1.260	1.446	1.575	1.843

	$\gamma = .975$						
n \ p	.75	.90	.95	.975	.99	.995	.999
2	32.342	45.303	53.306	60.320	68.524	74.127	85.705
3	4.426	6.263	7.412	8.424	9.613	10.428	12.115
4	2.178	3.096	3.673	4.183	4.785	5.198	6.055
5	1.489	2.120	2.519	2.872	3.290	3.577	4.173
6	1.169	1.667	1.982	2.262	2.593	2.821	3.294
7	0.986	1.407	1.674	1.911	2.192	2.385	2.788
8	0.867	1.237	1.473	1.682	1.930	2.101	2.457
9	0.783	1.119	1.332	1.521	1.746	1.901	2.224
10	0.722	1.031	1.227	1.402	1.610	1.753	2.051

Table 1 Tolerance Factors for Normal Distributions Utilizing the Range (continued)

	$\gamma = .99$						
n \ p	.75	.90	.95	.975	.99	.995	.999
2	80.867	113.274	133.283	150.821	171.332	185.343	214.292
3	7.059	9.988	11.820	13.434	15.331	16.630	19.320
4	2.982	4.237	5.026	5.725	6.549	7.114	8.287
5	1.903	2.710	3.219	3.671	4.205	4.572	5.334
6	1.433	2.042	2.428	2.771	3.177	3.456	4.037
7	1.176	1.678	1.996	2.279	2.614	2.845	3.326
8	1.015	1.449	1.725	1.970	2.260	2.460	2.877
9	0.904	1.290	1.536	1.755	2.014	2.193	2.565
10	0.823	1.176	1.400	1.600	1.837	2.000	2.341

	$\gamma = .995$						
n \ p	.75	.90	.95	.975	.99	.995	.999
2	161.736	226.552	266.572	301.648	342.671	370.693	428.592
3	9.935	14.058	16.635	18.907	21.577	23.405	27.191
4	3.773	5.361	6.360	7.244	8.286	9.002	10.485
5	2.276	3.241	3.850	4.391	5.029	5.468	6.380
6	1.662	2.369	2.817	3.215	3.685	4.009	4.683
7	1.338	1.910	2.272	2.594	2.975	3.238	3.784
8	1.135	1.620	1.928	2.202	2.527	2.751	3.217
9	1.003	1.432	1.704	1.947	2.235	2.433	2.847
10	0.905	1.292	1.539	1.758	2.018	2.198	2.572

Table 2. Tolerance Factors for Normal Distributions Utilizing the Average Range

Factors K for two-sided tolerance limits such that the probability is γ that at least P of the distribution will be included between $\bar{x} \pm K\bar{R}_{m,5}$, where $\bar{R}_{m,5}$ is the sample average range computed on the basis of m subgroups each of size 5.

	$\gamma = .75$						
m \ P	.75	.90	.95	.975	.99	.995	.999
2	.638	.911	1.085	1.240	1.423	1.550	1.814
3	.602	.860	1.024	1.171	1.345	1.465	1.716
4	.582	.832	0.992	1.134	1.303	1.419	1.663
5	.570	.816	0.972	1.111	1.276	1.391	1.630
6	.562	.803	0.957	1.095	1.258	1.371	1.606
7	.556	.794	0.946	1.082	1.244	1.355	1.588
8	.551	.787	0.938	1.073	1.233	1.343	1.575
9	.547	.782	0.932	1.065	1.224	1.334	1.564
10	.544	.777	0.926	1.059	1.217	1.326	1.555

	$\gamma = .90$						
m \ P	.75	.90	.95	.975	.99	.995	.999
2	.772	1.103	1.313	1.501	1.723	1.876	2.196
3	.699	0.999	1.190	1.361	1.563	1.703	1.994
4	.662	0.946	1.127	1.288	1.480	1.613	1.889
5	.638	0.913	1.087	1.243	1.428	1.556	1.823
6	.622	0.889	1.060	1.212	1.392	1.517	1.778
7	.610	0.872	1.039	1.188	1.365	1.488	1.744
8	.601	0.859	1.023	1.170	1.345	1.465	1.717
9	.593	0.848	1.010	1.156	1.328	1.447	1.696
10	.587	0.839	1.000	1.144	1.314	1.432	1.679

Table 2. Tolerance Factors for Normal Distributions Utilizing the Average Range (continued)

	$\gamma = .95$						
m \ P	.75	.90	.95	.975	.99	.995	.999
2	.875	1.249	1.487	1.700	1.951	2.125	2.486
3	.770	1.100	1.311	1.498	1.721	1.875	2.196
4	.717	1.025	1.221	1.396	1.604	1.748	2.048
5	.685	0.980	1.167	1.334	1.533	1.670	1.957
6	.663	0.947	1.129	1.291	1.483	1.616	1.894
7	.647	0.925	1.102	1.260	1.448	1.578	1.849
8	.634	0.907	1.080	1.235	1.419	1.547	1.813
9	.624	0.892	1.063	1.216	1.397	1.522	1.784
10	.616	0.880	1.049	1.200	1.378	1.502	1.761

	$\gamma = .975$						
m \ P	.75	.90	.95	.975	.99	.995	.999
2	.981	1.402	1.669	1.907	2.189	2.384	2.789
3	.840	1.201	1.430	1.635	1.878	2.046	2.396
4	.772	1.103	1.314	1.503	1.726	1.881	2.204
5	.730	1.045	1.244	1.422	1.634	1.780	2.086
6	.702	1.004	1.196	1.367	1.571	1.712	2.006
7	.681	0.974	1.160	1.327	1.525	1.661	1.947
8	.666	0.952	1.134	1.297	1.490	1.624	1.903
9	.653	0.933	1.112	1.272	1.461	1.592	1.866
10	.643	0.919	1.095	1.252	1.439	1.568	1.838

Table 2. Tolerance Factors for Normal Distributions Utilizing the Average Range (continued)

	$\gamma = .99$						
m \ p	.75	.90	.95	.975	.99	.995	.999
2	1.130	1.614	1.922	2.196	2.521	2.745	3.212
3	0.935	1.336	1.591	1.819	2.089	2.276	2.665
4	0.843	1.205	1.435	1.641	1.885	2.054	2.407
5	0.788	1.127	1.342	1.535	1.763	1.921	2.252
6	0.752	1.075	1.280	1.464	1.682	1.833	2.148
7	0.725	1.037	1.235	1.413	1.623	1.769	2.073
8	0.705	1.008	1.201	1.373	1.578	1.720	2.016
9	0.689	0.985	1.173	1.342	1.542	1.680	1.969
10	0.675	0.966	1.151	1.316	1.512	1.648	1.932

	$\gamma = .995$						
m \ p	.75	.90	.95	.975	.99	.995	.999
2	1.252	1.788	2.129	2.432	2.792	3.041	3.558
3	1.009	1.442	1.718	1.964	2.256	2.457	2.878
4	0.897	1.283	1.528	1.747	2.007	2.187	2.562
5	0.832	1.191	1.417	1.621	1.862	2.029	2.378
6	0.789	1.127	1.343	1.536	1.765	1.923	2.254
7	0.758	1.083	1.291	1.476	1.696	1.848	2.166
8	0.734	1.049	1.250	1.430	1.643	1.790	2.098
9	0.715	1.023	1.219	1.394	1.601	1.745	2.045
10	0.700	1.001	1.192	1.364	1.567	1.708	2.001

ON THE CHOICE OF SAMPLING INSPECTION PLANS

By Donald Guthrie

This paper summarizes a joint research effort of myself and Dr. M. Vernon Johns of Stanford University. The primary objective is basic research in sampling inspection schemes, but the method would be quite easily applicable, were the correct data available.

Suppose that we are presented with a lot consisting of N items, of which an unknown number D are defective. On the basis of observing d of these defectives in a sample of n we wish to decide whether to accept or reject the uninspected remainder of the lot. There are two problems, then, choosing the sample size n, and the acceptance number a.

The "classical" method of attacking such a problem is to choose an AQL or an AOQL or some similar index and consult the tabulated sampling plans to find a plan which matches these specifications. What we propose to do is to use cost considerations rather than AQL or AOQL to find the best n and a.

Actually it seems that cost considerations are often used in choosing an AQL but they are used subjectively, whereas we propose to use cost considerations objectively wherever possible. There will be cases arising in practice where the classical criteria will be more suitable, but this is a choice which must be made by the person choosing the plan at the outset.

Now, we adopt the following cost structure:

c_1 = Cost of accepting a defective item

c_2 = Cost of inspecting an item, that is, the time, etc., required to determine whether the item is defective or good.

c_3 = Additional inspection cost if the item is found defective, for example, the cost of replacing an item found defective with a good item.

c_4 = cost of rejecting an item.

There are two possible decisions which may be made at the conclusion of the sampling inspection, acceptance or rejection of the remainder of the lot. If we were to accept the lot we would incur a cost of c_1 for each of the D - d defective items left in the lot, but if we were to reject the lot we would incur a cost of c_4 for each of the N - n uninspected items remaining in the lot. Including the sampling cost, we may summarize the total cost by the following table:

Decision	Cost
Accept	$c_1 (D - d) + c_2 n + c_3 d$
Reject	$c_4 (N - n) + c_2 n + c_3 d$

Now suppose that we know the true value of p (we never do, of course). Then the expected costs would be as follows:

Decision	Expected Cost
Accept	$c_1 p(N - n) + c_2 n + c_3 np$
Reject	$c_4 (N - n) + c_2 n + c_3 np$

If we really knew p, we would minimize the expected cost by accepting whenever $p \leq c_4/c_1$ and rejecting whenever $p > c_4/c_1$. This then defines an "acceptable quality level" in the sense that $p \leq c_4/c_1$ is desirable and $p > c_4/c_1$ is undesirable under the above cost structure.

We make the further assumption that p varies from lot to lot. In this context p is not the proportion of defectives in the lot, but rather the probability that the machine which produces the items will produce any one item defectively, therefore is never actually observable. The variation of p from lot to lot is assumed to follow a known probability distribution F(p). That is, within each lot the probability an item is produced defectively is a constant, but between lots p is determined by independent observations on a population with cummulative distribution function F(p).

This situation could arise, for example, if different lots were produced on different days. It would be reasonable to suppose that p would change from day to day, and the nature of this variation may be described by F(p).

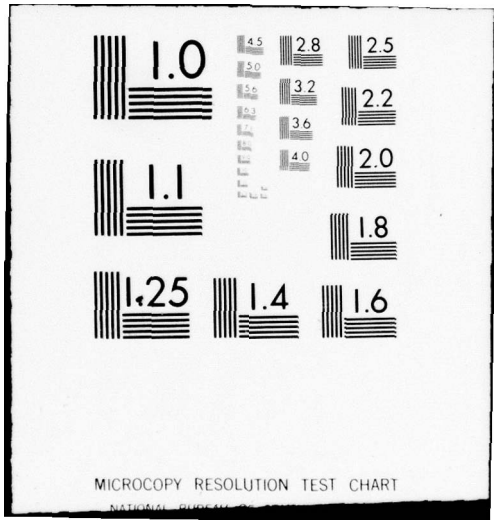
AD-A074 150 ARMY CHEMICAL CORPS ENGINEERING COMMAND ARMY CHEMICAL--ETC F/G 12/1
PROCEEDINGS OF THE ANNUAL STATISTICAL ENGINEERING SYMPOSIUM (3R--ETC(U)
MAY 57 D R HOWES

UNCLASSIFIED NL

2 OF 2
ADA 074150



END
DATE FILMED
10-79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART

Under this assumption, we may define a risk or total expected cost, with expectations taken from lot to lot as well as within lots. If we define

$$\delta(d) = \begin{cases} 1 & \text{if the lot is accepted} \\ 0 & \text{if the lot is rejected,} \end{cases}$$

then the total expected cost is

$$R_N(n) = E(c_2 n + c_3 d + c_1 \delta(d)\ (D - d) + c_4(1 - \delta(d))\ (N - n))$$

Some simple algebraic calculations lead to showing at this point that the best $\delta(d)$ (the one which minimizes $R_N(n)$ for fixed n) is

$$\delta^*(d) = \begin{cases} E(p|d) \leq c_4/c \\ E(p|d) > c_4/c_1 \end{cases}$$

where $E(p|d)$ represents the conditional expected value of p (with respect to the distribution $F(p)$) given that d defectives have been observed in a sample of n. Now, clearly $E(p|d)$ is an increasing function of d so we may equivalently express $\delta(d)$ by

$$\delta^*(d) = \begin{cases} 1 & d \leq [b(n)] \\ 0 & d > [b(n)] \end{cases}$$

where $b(n)$ is determined by solving $E(p|b(n)) = c_4/c_1$. That is, we accept whenever $d \leq [b(n)]$ and reject whenever $d > [(b(n)]$. In the more involved mathematical parts of this paper, we have derived asymptotic expressions for $b(n)$ and have used this in calculating the best sample size $n(N)$ as a function of the lot size.

There are two cases considered: (1) $F(p)$ has a density which is continuously differentiable at c_4/c_1, and (2) $F(p)$ has probability on only two points, one on each side of c_4/c_1. The results are as follows:

Case	Best acceptance number	best sample size
1	$b(n) = (c_4/c_1)\ n+c+o(1)$	$n(N) = AN^{1/2} + o(N^{1/2})$
2	$b(n) = k_o+k_1 n+o(1)$	$n(N) = B\log N + o(\log N)$

The numbers c, k_0, k_1, A, and B all depend on the various costs and the nature of F(p). The symbol o(g(N)) indicates that that quantity goes to zero if divided by g(N) as $N \to \infty$. Similarly o(1) means that that term goes to zero as $n \to \infty$.

In conclusion, there are two possible gerneralizations of the problem which should be exposed. The first and most important is that an item may be classified by a value of a random variable X associated with it. In the case just considered X takes on the values zero or one depending on whether the item is good or defective. In more general cases X may be the number of defects per item or the amount by which a dimension differs from some nominal dimension. The solution given above carries over to a fairly wide class of distributions including most of those found in practice, e.g., Poisson, geometric, chi-square, etc. That is, if X is a random variable with one of these distributions, then the optimal sample size is proportional to the square root or the logarithm of the lot size, depending on which type of lot to lot distribution of the parameter is assumed to exist.

A second generalization concerns only the case where each item is defective or good. We might decide that there should be a cost for rejecting items only if they are good. That is c_4 might be replaced by c_4', the cost of rejecting a good item, and there would be no cost attached to rejecting a defective. This would change the total cost of a decision to the following:

Decision	Cost
Accept	$c_1(D - d) + c_2 n + c_3 d$
Reject	$c_4' \left[(N - D) - (n - d)\right] + c_2 n + c_3 d$

By application of the same analysis we determine that the optimal sample size is still proportional to the square root or logarithm of the lot size and the rejection number is proportional to the sample size, although the constants are different.

TIGHTENED MULTI-LEVEL CONTINUOUS SAMPLING PLANS *

C. Derman,[1] S. Littauer[2,3] and H. Solomon[2,3]
Columbia University

1. Introduction. Industrial needs have provoked some recent studies on continuous sampling. This procedure is especially of interest when the formation of inspection lots for lot-by-lot acceptance may be impractical or artificial as in conveyor-line production, or when there is an important need for rectifying quality of product as it is manufactured.

These newer papers are best considered in the light of the earlier papers of Dodge (3) and Wald and Wolfowitz (11). One point of departure from the Dodge type of plan has been the introduction of several levels of partial inspection with different rates of sampling in each level. Multi-level continuous sampling plans (which reduce to the Dodge plan when only one sampling level is tolerated) have been considered by Greenwood (8), Lieberman and Solomon (9), and Resnikoff (10). A plan based on the Wald-Wolfowitz approach, a scheme essentially handled by the methodology of sequential analysis, was created and developed by Girshick about 1948 in connection with a Census Bureau problem and has only recently been reported (7). The reader is referred to Bowker (1) for a more thorough account of continuous sampling plans.

The multi-level plan given in (9), namely MLP, allows for any number of sampling levels, subject to the provision that transitions can only occur between adjacent levels. Three generalizations of MLP, accomplished by altering the manner in which transition can occur, are analyzed in this paper. In each situation, we will make it more difficult to get to infrequent inspection than in MLP, and thus we can label these three plans as tightened plans. These three plans which will now be specifically defined obviously relate to more realistic situations for control of industrial processes. The three plans are given in language which assumes some familiarity with MLP, which is given in detail in (9).

---

* This paper is reproduced with the permission of The Annals of Mathematical Statistics.
Received May 24, 1956; revised September 14, 1956.
[1] Work sponsored in part by the Office of Scientific Research, U. S. Air Force.
[2] Work sponsored in part by the Office of Naval Research.
[3] Work sponsored in part by the Higgins Fund, Columbia University.

(a) The MLP-r x 1 Plan. We say we are in the jth sampling level if every $(1/f)^j$-th item produced is systematically sampled. If i consecutively inspected items are found clear of defects when sampling at the jth level, begin sampling at the (j + 1)-th level. On the other hand, if a defective item is found before this is accomplished, revert immediately to the (j - r)-th level, if $j > r$, or to the zero level, that is, one hundred percent inspection if $j \leq r$. Let inspection begin at the zero level. When r = 1, we have the MLP plan described in (9).

(b) The MLP-T Plan. This is exactly the same as the MLP-r x 1 Plan, except that when a defective is encountered, we immediately revert to one hundred percent inspection. This is obviously the tightest of the three multi-level plans considered in this paper and thus bears the label MLP-T.

(c) The MLP-r x s Plan. This plan follows exactly the same pattern as the MLP-r x 1, except that when i consecutively inspected items are found nondefective while on the jth sampling level, systematic sampling begins at level (j + s). We shall consider the case $r > s$, since we are concerned only with tightened multi-level plans. If r = s, we are effectively using the MLP Plan.

2. Summary. Each of these generalizations can be appraised under the assumption of an infinite number of sampling levels or a finite number, k, of sampling levels. Under the assumption of an infinite number of allowable sampling levels, it is possible to obtain explicit relationships between the AOQL and the parameters of the plan for MLP-r x 1 and MLP-T. Thus it is possible to graph contours of equal AOQL for each of these plans under these conditions. Approximations for contours of equal AOQL for the MLP-r x s Plan are then easily obtained. This makes feasible the possibility of a catalogue of continuous sampling plans which contains plans having a prescribed AOQL and thus aids immeasurably in the choice of an appropriate plan. As is demonstrated in the next sections, the following results are obtained, assuming that the production process is in statistical control and items found defective on inspection are replaced with good items. For the MLP-r x 1 Plan:

$$\text{(2.1)} \qquad \text{AOQL} = 1 - \left(\frac{f - f^{r+1}}{1 - f^{r+1}}\right)^{1/i} .$$

When r = 1, this reduces to the result previously obtained in (9). For the MLP-T Plan:

$$\text{(2.2)} \qquad \text{AOQL} = 1 - f^{1/i}$$

This result can also be obtained heuristically by letting r approach infinity in MLP-r x 1. For the MLP-r x s Plan (r > s) bounds and sometimes exact AOQL's can be obtained using the previous two results. For example, if r = 4 and s = 2 and f is given, the MLP-2 x 1 Plan for $f' = f^2$ will be the same plan and hence have the same AOQL. More generally for a given f we can write

$$\text{(2.3)} \qquad \text{AOQL}_{r'' \text{ xs}} < \text{AOQL}_{\text{rxs}} < \text{AOQL}_{r' \text{ xs}}$$

where r' = greatest number less than r that is a multiple of s, and r" is the smallest number greater than r that is a multiple of s. For, if r' < r", the plan associated with r" is tighter and the added protection thus insures a better outgoing quality, i.e., a smaller AOQL. Under the assumption of a finite number, k, of allowable sampling levels, the AOQ function for MLP-T is obtained, and it is seen that the use of digital computers may be expedient for the computation of AOQL contours. This was exactly the situation, for finite levels, in (9). The main results of the paper are obtained through the use of Markov chain techniques which are developed in Section 3. In these plans inspection, as described, is by systematic sampling. However, the AOQ and AOQL results also hold when inspection in each level is accomplished by random sampling —i.e., in the kth level, each item in the block of f^{-k} items has probability f^k of being chosen for inspection.

3. Markov Chain Result. Let $\{X_n\}$ (n = 0, 1, ...) denote an irreducible recurrent positive Markov chain with states $\{E_j\}$ (j = 0, 1, ...). Let $\{p_{ij}\}$ (i, j = 0, 1, ...) denote the probability of transition from state E_i to E_j. It is known (see (5)), that a unique sequence $\{v_i\}$ exists such that

$$\text{(3.1)} \qquad \begin{aligned} &\sum_{i=0}^{\infty} v_i p_{ij} = v_j, && (j = 0, 1, \ldots), \\ &v_i > 0, && (i = 0, 1, \ldots), \\ &\sum_{i=0}^{\infty} v_i = 1. \end{aligned}$$

The v_i's are sometimes referred to as "steady state" probabilities.

Now let $A = \{E_{ji}\}$ be a subset of the states. Let $Y_0, Y_1, \ldots$ be successive members of $\{X_n\}$ which take on values in A. Since the chain is recurrent, infinitely many such Y's will exist with probability one. It was shown by Derman (2) that $\{Y_k\}$ $(k = 0, 1, \ldots)$ is also a Markov chain; and if $\{p'_{ij}\}$ $(E_i, E_j \in A)$ are its transition probabilities, then the solutions v'_j of

$$\sum_{E_i \in A} v'_i p'_{ij} = v'_i \qquad (E_j \in A),$$

$$v'_i > 0 \qquad (E_i \in A), \tag{3.2}$$

$$\sum_{E_i \in A} v'_i = 1$$

are given by

$$v'_i = \frac{v_i}{\sum_{E_j \in A} v_j} \qquad (E_i \in A). \tag{3.3}$$

Suppose $A_1 = \{E_j\}$ $(j = 1, 2, \ldots)$; $A_2 = \{E_j\}$ $(j = 2, 3, \ldots)$; $\ldots A_g = \{E_j\}$ $(j = g, g + 1, \ldots)\ldots$ are subsets to be considered. Let $\{Y_k(g)\}$ denote the Markov chain defined over A_g. Also let $E_j(g)$ $(j = 0, 1, \ldots)$, the states for the chain $\{Y_k(g)\}$, be a relabeling of the states $E_k (k = g, \ldots)$ by letting $j = k - g$. Finally let $p_{ij}(g)$ denote the probability of transition from state $E_i(g)$ to state $E_j(g)$ in the chain $\{Y_k(g)\}$. Our main tool is the following theorem

THEOREM. If $p_{ij} = p_{ij}(g)$ $(i,j = 0, \ldots ; g = 1, \ldots)$, then

$$v_j = v_0(1 - v_0)^j \qquad (j = 1, \ldots). \tag{3.4}$$

PROOF. Let $\{v_j(g)\}$ denote the solution of (3.1) for the chain $\{Y_k(g)\}$. Since the transition probabilities, by hypothesis, are the same regardless of which chain is under consideration, $v_i(g) = v_i$ $(i = 0, 1, \ldots)$. However, from (3.3) we have

$$(3.5) \qquad v_0 = v_0(g) = \frac{v_g}{\sum_{j=g}^{\infty} v_j} = \frac{v_g}{1 - \sum_{j=0}^{g-1} v_j} \qquad (g = 1,2,\ldots).$$

Thus by induction,

$$v_j = v_0(1 - v_0 - \ldots - v_{j-1})$$

$$(3.6) \qquad = v_0\left[1 - v_0 - v_0 \sum_{i=1}^{j-1} (1 - v_0)^i\right]$$

$$= v_0(1 - v_0)^j \qquad (j = 1, \ldots),$$

and the theorem is proved.

We shall apply the theorem in the following case. Suppose

$$p_{i,i+1} = \alpha > 0 \qquad (i = 0,1, \ldots),$$

$$p_{i,0} = 1 - \alpha \qquad (i = 0,1, \ldots,r),$$

$$p_{i,i-r} = 1 - \alpha \qquad (i > r).$$

It is clear that the chain is irreducible. It also follows from a slightly modified theorem of Foster ((6), Theorem 5, p. 81) that the chain is recurrent positive if $\alpha < r/(r + 1)$. Intuitively this condition guarantees a sufficient pull to the left, thereby insuring the existence of the steady-state probabilities inherent in a recurrent positive chain. Furthermore, it is easily seen that the conditions of the theorem are satisfied so that the v_j have the form (3.4). From (3.1), $j = 0, v_0$ is determined by the following equation

$$(3.7) \qquad (1 - \alpha)\left\{\frac{1 - (1 - v_0)^{r+1}}{v_0}\right\} = 1,$$

and thus any v_j can be obtained.

4. Application to MLP-r x 1 infinite-level plan. The multilevel plans can now be studied from the point of view of a Markov chain $\{X_n\}$ and the results in Section 3 employed. We let E_{jm} $(j = 0, 1, \ldots; m = 0, \ldots, i - 1)$ denote the state of such a chain where we say that X_n is in state E_{jm} if just after the nth item has been inspected, the process is in the jth sampling level (i.e., every (f^{-j})th item inspected) and m nondefectives have been observed successively while in the jth level. Suppose the process is in a state of control such that p is the probability of a defective being produced. The transition probabilities are then given by

$$P(E_{jm} \rightarrow E_{j,m+1}) = 1 - p = q \qquad (j = 0, 1, \ldots; m = 0, 1, \ldots, i - 2)$$

$$(4.1) \qquad \begin{aligned} P(E_{j,i-1} \rightarrow E_{j+1,0}) &= q && (j = 0, 1, \ldots), \\ P(E_{jm} \rightarrow E_{j-r,0}) &= p && (j = r, \ldots), \\ P(E_{jm} \rightarrow E_{00}) &= p && (j = 1, \ldots, r-1), \end{aligned}$$

The chain is easily seen to be irreducible. From Foster's theorem it is seen to be recurrent positive if $q^i < r/(r + 1)$. We shall assume $q^i < r/(r + 1)$ for the present. Now let $A = \{E_{j0}\}$ be a subset of the states and let $\{Y_k\}$ denote the chain defined over it. The chain is of the form of the special case considered in section 3 with $\alpha = q^i$. Let $\{v'_j\}$ and $\{v_{jm}\}$ denote the steady-state probabilities of the chains $\{Y_k\}$ and $\{X_n\}$, respectively. Using (3.1), (3.5) and (4.1) it follows that

$$(4.2) \qquad v_{jm} = \frac{1 - q}{1 - q^i} v'_j \, q^m \qquad (m = 0, 1, \ldots, i - 1; \; j = 0, 1, \ldots).$$

For from (3.1)

$$v_{jm} = v_{j0} q^m \qquad (m = 0, \ldots, i - 1; \; j = 0, 1, \ldots),$$

and from (3.5)

$$v'_j = \frac{v_{j0}}{\sum_{k=0}^{\infty} v_{k0}} \qquad (j = 0, 1, \ldots).$$

Hence,

$$v_{jm} = \sum_{k=0}^{\infty} v_{k0} v'_j q^m \qquad (j = 0, 1, \ldots);$$

but summing over j and m we get, since $\Sigma_{j,m} v_{jm} = 1$,

$$\sum_{k=0} v_{k0} = \frac{1 - q}{1 - q^i}.$$

From (4.2) it is clear that v'_j is the sum of the steady-state probabilities of being in the jth level of sampling. Also from (3.4)

$$v'_j = v'_0 (1 - v'_0)^j \qquad (j = 1, 2, \ldots), \tag{4.3}$$

where v'_0 is given by (3.7) with $\alpha = q^i$; namely,

$$(1 - q^i)\left[\frac{1 - (1 - v'_0)^{r+1}}{v'_0}\right] = 1,$$

where as previously remarked, v'_0 is the probability of being in one hundred percent inspection.

Now that we have expressions for the steady-state probabilities, we proceed with the derivation of the AOQ functions and the AOQL. Let $h(X_n) = f^{-j}$ for $X_n = E_{jm}$. It is easily verified that the reciprocal of the average fraction inspected after n inspections is

$$F_n^{-1} = \frac{1}{n} \sum_{v=1}^{n} h(X_v). \tag{4.4}$$

It follows from the Birkhoff ergodic theorem, applicable for stationary Markov chains of the type considered here (see Doob (1), p. 460), that

$$(4.5) \qquad F^{-1} = \lim_{n\to\infty} F_n^{-1} = \sum_{j=0}^{\infty} f^{-j} \sum_{m=0}^{i-1} v_{jm}.$$

Now F^{-1} denotes the reciprocal of the average fraction inspected for all sequences (except for a set having probability 0); for let $t_k = \sum_{m=1}^{k} h(X_m)$ = number of items produced during the first k inspections. Formula (4.5) says that $k/t_k \to F$ as $k \to \infty$. Let $t_k < t < t_{k+1}$. Then since k = number of items inspected in the first t items produced, the inequalities

$$t\frac{k}{t_{k+1}} < \frac{k}{t} \leqq \frac{k}{t_k}$$

imply that $\lim_{k\to\infty} k/t \to F$ with probability 1.

If $q^i \geqq r/(r+1)$, it can be shown more directly that $F^{-1} = \infty$ with probability 1. If v_0' exists and is positive, it follows from the theory of recurrent Markov chains that $q^i < r/(r+1)$. Thus since $0<f<1$, we have from (4.2), (4.3), (4.5) and the last remark that

$$(4.6) \qquad F^{-1} = v_0' \left(\frac{1}{1 - \frac{1 - v_0'}{f}} \right), \quad \text{when } (f > 1 - v_0'),$$

$$= \infty, \qquad \text{otherwise.}$$

Hence since it can easily be shown that AOQ = p(1 - F), we have

$$(4.7) \qquad AOQ = (1-q)\left(\frac{1-f}{f}\right)\frac{1-v_0'}{v_0'}, \quad \text{when } (f > 1 - v_0'),$$

$$= 1 - q, \qquad \text{otherwise.}$$

Now suppose it is true that the AOQ is an increasing function of q as long as $f > 1 - v_0'$. Then from (4.7) it would follow that

$$\text{AOQL} = 1 - q_0, \tag{4.8}$$

where q_0 is the value of q such that $f = 1 - v_0'$. From (3.7) with $\alpha = q^i$, it is easily established that

$$q_0 = \left(\frac{f - f^{r+1}}{1 - f^{r+1}}\right)^{1/i},$$

so that

$$\text{AOQL} = 1 - \left(\frac{f - f^{r+1}}{1 - f^{r+1}}\right)^{1/i} \tag{4.9}$$

We now show that the AOQ is an increasing function of q as long as

$$q < \left(\frac{f - f^{r+1}}{1 - f^{r+1}}\right)^{1/i} \qquad (\text{i.e., } f > 1 - v_0').$$

Let

$$\phi(q) = \left(\frac{f}{1 - f}\right) \quad \text{AOQ} = (1 - q)\,\frac{1 - v_0'}{v_0'}$$

and

$$V(q) = \frac{1 - v_0'}{v_0'}.$$

Then

$$\frac{d\phi(q)}{dq} = -V(q) + (1 - q)\,\frac{dV(q)}{dq}. \tag{4.10}$$

It is necessary to show that the right-hand side of (4.10) is positive or

$$\frac{V(q)}{(1 - q)\frac{dV(q)}{dq}} \leq 1. \tag{4.11}$$

But, using (3.7) with $\alpha = q^i$,

$$(4.12)\quad \frac{dV(q)}{dq} = \left(-\frac{1}{v_0'^2}\right) \left\{ \frac{iq^{i-1}v_0'}{(1-q^i)^2 \left[(r+1)(1-v_0')^r - \frac{1}{(1-q^i)}\right]} \right\}.$$

Thus the left side of (4.11) becomes

$$(4.13)\quad \frac{-(1-q^i)\left[(r+1)(1-v_0')^{r+1}(1-q^i) - (1-v_0')\right]}{iq^{i-1}(1-q)}.$$

From (3.7) it follows that $(1-v_0')^{r+1} = \left[(1-v_0')-q^i\right]/(1-q^i)$. Hence (4.13) becomes

$$(4.14)\quad -\frac{q}{i}\left(\frac{1-q^i}{1-q}\right)\left[\frac{(1-v_0')^r}{q^i} - (r+1)\right].$$

But from (3.7)

$$q^i = (1-v_0')\,\frac{1-(1-v_0')^r}{1-(1-v_0')^{r+1}} \leq 1 - v_0'.$$

Hence

$$\frac{(1-v_0')^r}{q^i} \geq r,$$

and the smallest value over the range $f > 1 - v_0'$ which the bracket factor in (4.14) can take is minus one. Thus the largest value that (4.14) can reach is

$$(4.15)\qquad \frac{(1-q^i)}{1-q}\left(\frac{q}{i}\right).$$

But

$$\frac{1-q^i}{1-q}\left(\frac{q}{i}\right) \qquad \frac{q+q^2+\cdots+q^i}{i} < 1.$$

This proves (4.11).

5. The MLP-T Plan. We consider first an infinite number of sampling levels. Let E_{jm} be as in the previous section The transition probabilities are now

$$P(E_{jm} \rightarrow E_{j,m+1}) = q$$

$$(j = 0, 1, \ldots;\ 0 < m \leqq i - 2),$$

$$P(E_{j,i-1} \rightarrow E_{j+1,0}) = q \qquad (j = 0, 1, \ldots),$$

$$P(E_{jm} \rightarrow E_{00}) = 1 - q \qquad (\text{for all } j, m).$$

Of course, $0 < q < 1$.

It can be shown in this case that

$$v_{jm} = pq^{ji+m}$$

$$(j = 0, 1, \ldots;\ m = 0, \ldots, i-1),$$

and as before that

$$F^{-1} = \sum_{jm} f^{-j} v_{jm} = \frac{1 - q^i}{1 - \frac{q^i}{f}} \qquad (f > q^i),$$

$$= \infty \qquad (f \leqq q^i);$$

and

$$AOQ = \frac{(1 - q)\, q^i}{1 - q^i} \left(\frac{1 - f}{f} \right) \qquad (f > q^i),$$

$$= 1 - q \qquad (f \leqq q^i).$$

It can easily be shown that AOQ is an increasing function of q for $0 < q^i < f$.

Hence

$$AOQL = 1 - f^{1/i}.$$

Now let the number of sampling levels, k, be finite. For this case we need only modify the function $h(X_n)$ such that

$$h(X_n) = f^{-j} \quad \text{when} \quad X_n = E_{jm} \qquad (j \leqq k),$$

$$= f^{-k} \quad \text{when} \quad X_n = E_{jm} \qquad (j > k),$$

where here we persist with the notation E_{jm} as if the $k = \infty$ plans are in effect. In similar fashion we have

$$F^{-1} = p \sum_{j=0}^{k-1} \sum_{m=0}^{i-1} f^{-j} q^{ji+m} + p \sum_{j=k}^{\infty} \sum_{m=0}^{i-1} f^{-k} q^{ji+m}$$

$$= (1 - q^i) \frac{1 - (q^i/f)^k}{1 - q^i/f} + (q^i/f)^k.$$

For $k = 1$, we have the Dodge Plan, and get the following result as in (3):

$$F^{-1} = \frac{f}{f + q^i(1 - f)}.$$

For $k = 2$,

$$F^{-1} = 1 + q^i\left(\frac{1 - f}{f}\right) + q^{2i}\left(\frac{1 - f}{f^2}\right):$$

In order to obtain AOQL contours for this situation, as for higher values of k, the use of digital computers would be expedient.

REFERENCES

(1) A. H. Bowker, "A survey of continuous sampling plans," Proceedings of the Third-Berkeley Symposium on Mathematical Statistics and Probability, Vol. V, University of California Press, 1956, pp. 75-86.

(2) C. Derman, "Some contributions to the theory of denumerable Markov chains," Trans. Amer. Math. Soc, Vol. 79, No. 2 (1955), pp. 541-555.

(3) H. F. Dodge, "A sampling inspection plan for continuous production," Ann. Math. Stat., Vol. 14 (1943), pp. 264-279.

(4) J. L. Doob, Stochastic Processes, John Wiley and Sons 1953.

(5) W. Feller, Probability Theory and Its Applications, John Wiley and Sons, 1950.

(6) F. G. Foster, "Markoff chains with an enumerable number of states and a class of cascade processes," Proc. Cambridge Phil. Soc., Vol. 47 (1951), pp. 77-85.

(7) M. A. Girshick, "A sequential inspection plan for quality control," Technical Report No. 16, Applied Mathematics and Statistics Laboratory, Stanford University, 1954.

(8) J. A. Greenwood, "A continuous sampling plan and its operating characteristics," Bureau of Ordnance, Navy Dept., Washington, D. C. unpublished memorandum.

(9) G. Lieberman and H. Solomon, "Multi-level continuous sampling plans", Ann. Math. Stat., Vol. (26), No. 4 (1955), pp. 686-704.

(10) G. Resnikoff, "Some modifications of the Lieberman-Solomon multi-level continuous sampling plan, MLP," Technical Report No. 26, Applied Mathematics and Statistics Laboratory, Stanford University, 1956.

(11) A. Wald and J. Wolfowitz, "Sampling inspection plans for continuous production which insure a prescribed limit on the outgoing quality," Ann. Math. Stat., Vol. 16 (1945), pp. 30-49.

TIME AS A RESPONSE

By G. Stanley Woodson
Chief, Biostatistics Section
Medical Research Directorate
U. S. Army Chemical Warfare Laboratories

INTRODUCTION

Time, as a response, enters into many investigations. It may appear as a part of the treatment, or as part of the response, or as both at the same time. On other occasions, when time is measured from the point in the proceedings when a treatment is applied, interest is centered upon the length of time until a response occurs. It is with this last situation that we shall concern ourselves.

There arise occasions when interest is centered in describing, over the complete range of treatment levels, the time of occurrence of some more or less well defined end-point; the time to rupture of metals under stress, the latent period of the action of an injected drug, the length of time required to achieve a certain level of activity, or almost any situation where we have need to determine the speed of response. In general it has been noted that the functional form of such a relationship between time and treatment level is reasonably well described by an hyperbola. In some cases, however, the hyperbolic form seems to hold true only at the higher levels of treatment. Here, the lower level treatments result in a decreasing function with decreasing level of treatment. This sort of function occurs in particular when we plot the mean time of response against the level of treatment, ignoring the non-responses.

THE PROBLEM

During experiments at the Medical Research Directorate, U. S. Army Chemical Warfare Laboratories, the problem arose of describing the time necessary for certain agents to have an effect. This problem has, of course, wide practical application in chemical warfare. We were immediately faced with the question of defining an expected time of response in a group where we expect only a partial response. By a partial response, I refer to the case where only a portion of a given group reacts. For instance, if only five out of a group react, just what do we mean when we say "mean reaction time"?

Obviously we need some measure which will not only give the expected response in terms of numbers of minutes or hours, but one that will also tell how many responses to expect. In the past the general practice has been to deal with the reciprocal of the time (2, 3, 4) with the view in mind that late responses would carry an appropriately small weight and each non-response could be included by adding nothing to the total and adding one to the number by which the total is divided to obtain the mean response time. Some authors have utilized this approach and ignored the resulting small bend at the lower end of the curve (4,5). This approach does very nicely if we only desire to estimate the time of response, while making the tacit assumption that the proportion responding is 100 percent. Since, for our purposes, we cannot make this assumption, another approach is needed.

THE SOLUTION

The data which are to be described are shown in Figure I.[1] The simplest mathematical model which presents itself is one which states that the time to incapacitation is inversely proportional to the dose used. That is

$$T = B/k \qquad (1)$$

where T = time to response, measured from the instant the treatment has been applied.

B = a constant to be estimated

k = dose

It is immediately apparent that equation (1) does not present the true picture, for as dose becomes increasingly large, equation (1) tells us that T approaches zero, while the data approach some number greater than zero. That this happens is not too surprising in this case, since, regardless of the size of the dose administered, a certain minimal time is required for the agent to penetrate the tissues and start to have an effect. Therefore, consider

[1]The data used in this paper have been transformed in such a manner as to conceal the identity of the agent used without altering any of the statistical properties of the data.

the following modification of equation (1):

$$T = T_o + (B'/k) \qquad (2)$$

where T_o and B' are constants to be estimated.

Equation (2) says that the difference between this minimal time to start the toxicological process and the actual time to response is what is to be considered as being inversely proportional to the dose.

Now, notice that the times to respond tend to increase with decreasing dose until a particular area is reached, then they seem to stabilize. This would indicate that our model requires further modification.

In studying time responses of the type used here, it is usually noted that if only a partial response is obtained, then the responses that are noted seem to occur at fairly early times. That is, if a response occurs in connection with a low level treatment, then it generally occurs fairly quickly. It would seem that the weaker animals and materials respond first, the stronger ones either hold out longer or do not show a response at all.

You will notice that I stated "if only a partial response is obtained...". This would imply that the time of response is conditionally dependent upon the probability of obtaining a response. This seems only logical. If no response occurs, then we certainly can't measure the time of response.

A common mathematical expression (used in probit analysis) representing the proportion affected at various dose levels is

$$P = (1/\sqrt{2\pi}) \int_{-\infty}^{D + Ek} e^{-z^2/2} dz \qquad (3)$$

where P = expected proportion affected

D and E are constants to be estimated.

Making the assumption that equation (3) actually provides us with an estimate of the probability of the occurrence of a response in response to a dose (k) then we may incorporate

equations (2) and (3) to obtain the following rather formidable appearing mathematical model:

$$T = \left[T_o + (B'/k)\right] \; (1/\sqrt{2\pi}) \int_{-\infty}^{D + Ek} e^{-z^2/2} dz. \qquad (4)$$

This model gives us a predicted response time which is the mean expected response time of those animals that actually respond. Fitting this equation to the data results in the function shown in Figure 2. In fitting this function to the data, the sum of squares of the differences between the logarithms (7) was minimized. This sum of squares was considered rather than the sum of the squares of the arithmetic differences because this transform resulted in a greater degree of homoscedasticity about the line.

Thus we have arrived at a solution to our original problem, which was to describe the time necessary for the agent to have an effect.

THE EXPANDED PROBLEM

In order to put our solution to work in practical situations, something more than the solution found here is needed. When we ask such a question as "if I use a dose of k_i, how long will it be before 25 percent of the exposed group react? 50 percent? 75 percent? 95 percent?" the inadequacies of the present model become apparent. Such information simply cannot be obtained from it.

Fortunately, thanks to the work of Shewhart (8), Bowker (1), Wald (9, 10, 11), and others (6, 12, 13), a method is available to us which allows us to make some relatively simple adjustments to our model and obtain the sort of answers we need. Consider that portion of the curve where a 100 percent response is indicated. Here, we have a series of observations distributed around a linear regression, and it is a relatively simple matter to define 'tolerance' limits around the regression line that demarcate upper limits for any desired proportion. If we attempt to do this at the lower levels of treatment we run into difficulties, for here we are no longer talking about the exposed population, but instead we are describing only the responding portion of the exposed animals.

Referring back to equation (4), if we select a dose low enough to affect only a portion of the exposed group, we can predict the proportion expected to respond at that dose as well as the mean time of response within the responding group. If we now define tolerance limits around this mean time, we see that we are actually defining some unknown proportion of the total exposed group. It is a relatively simple matter to find out what this proportion is. Let us choose a dose which achieves a 50 percent response and mark off a time limit by which we expect 75 percent of those responding to have responded. It is immediately obvious that by this same time limit we have defined a limit such that 75 percent of 50 percent, or 37 1/2 percent of the exposed population is expected to have responded. To generalize on this, we may say that the proportion of the exposed population which lies below a particular time limit is described by the product of the proportion responding and the proportion of those responding which lie below the same limit.

THE FINAL SOLUTION

Putting all these ideas together, we have constructed a graph (Figure 3) giving the time limits by which a particular portion of the exposed population will react over the whole range of dose levels considered. To use this graph, we select some dose, say 3000, and read upwards. We note that:

10 percent are expected to respond by a time of 13

30 percent are expected to respond by a time of 17

50 percent are expected to respond by a time of 20

70 percent are expected to respond by a time of 24

90 percent are expected to respond by a time of 32

95 percent are expected to respond by a time of 36

99 percent are expected to respond by a time of 47

Notice that if we choose a dose at which a partial response is expected, we find a slight difference in reading the graph. Say we had selected a dose of 500. At this dose we would read the following:

10 percent are expected to respond by a time of 35

30 percent are expected to respond by a time of 45

but we cannot find a line corresponding to 50 percent or to any proportion higher than 30 percent. This means that we do not expect to achieve a response rate of 50 percent; we only expect a response of something more than 30 percent but less than 50 percent at this particular dose.

SUMMARY

Based on the viewpoint that the time of response is conditionally dependent upon the probability of obtaining a response, a relatively simple method has been developed for determining time limits by which a certain proportion of a population exposed to a particular treatment is expected to respond. A graph has been constructed and presented which allows immediate and direct determination of these limits.

BIBLIOGRAPHY

1. Bowker, A. H., "Computation of Factors for Tolerance Limits on a Normal Distribution When the Sample is Large", Annals of Math. Stat., Vol. 17 (1946), pp. 238-240

2. Box, G. E. P., & Cullumbine, H., "The Relationship Between Survival Time and Dosage with Certain Toxic Agents", Brit. J. Pharmacol., Vol. 2 (1947), pp. 27-37.

3. Bryan, W. R., "Quantitative Studies of the Latent Period of Tumors Induced with Subcutaneous Injections of the Agent of Chicken Tumor. I. Curve Relating Dosage of Agent and Chicken Response", J. Nat. Cancer Res. Inst., Vol. 6 (1946), pp. 225-237.

4. Gaddum J. H., "Bioassays and Mathematics", Pharmacol. Rev., Vol. 5 (1953), pp. 87-134.

5. Gard, S., "Encephalomyelitis of Mice. II. A Method for the Measurement of Virus Activity", J. Exper. Med., Vol. 72 (1940), pp. 69-77.

6. Scheffe, H., & Tukey, J. W., "A Formula for Sample Sizes for Population Tolerance Limits", Annals Math. Stat., Vol. 15 (1944), p. 217.

Figure 1

RESPONSE TIMES

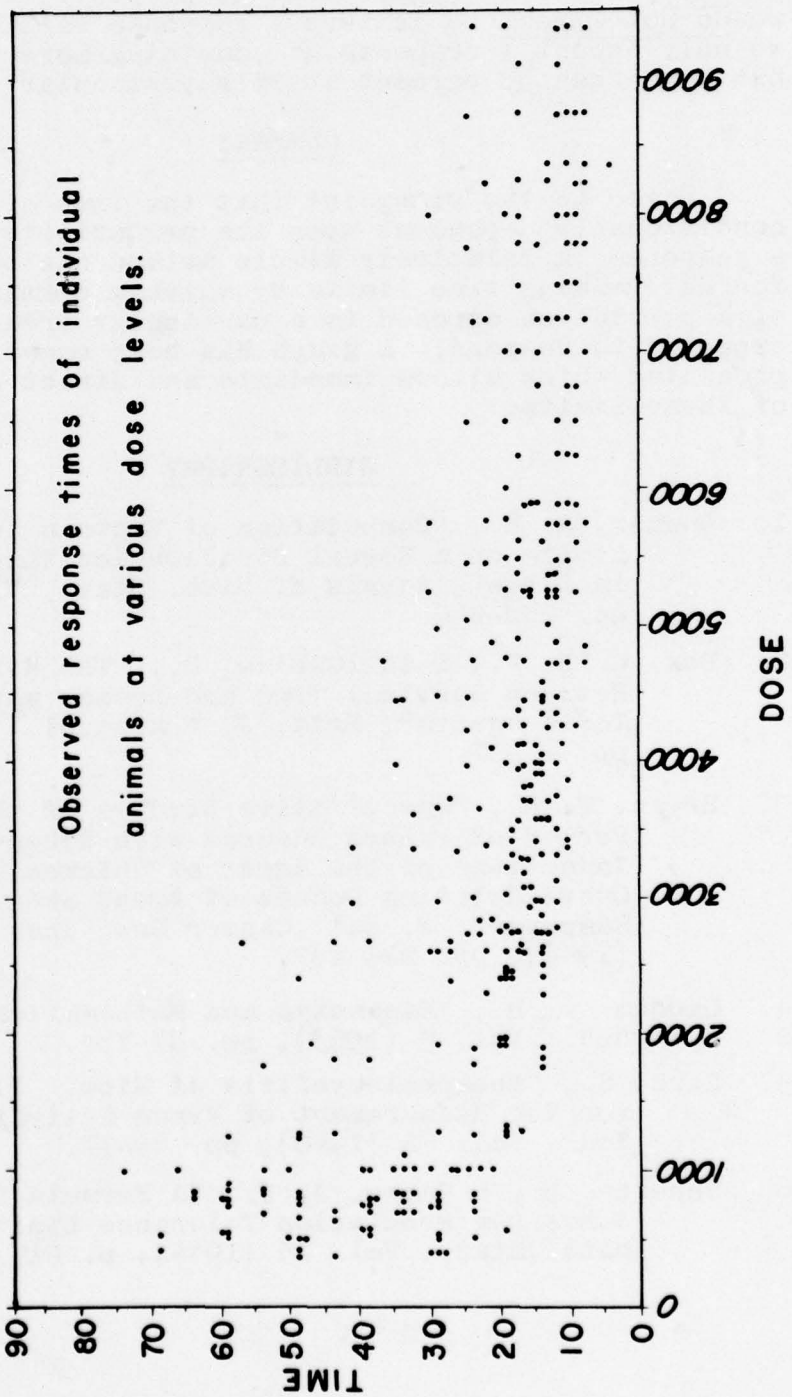


Figure 2

EXPECTED RESPONSE

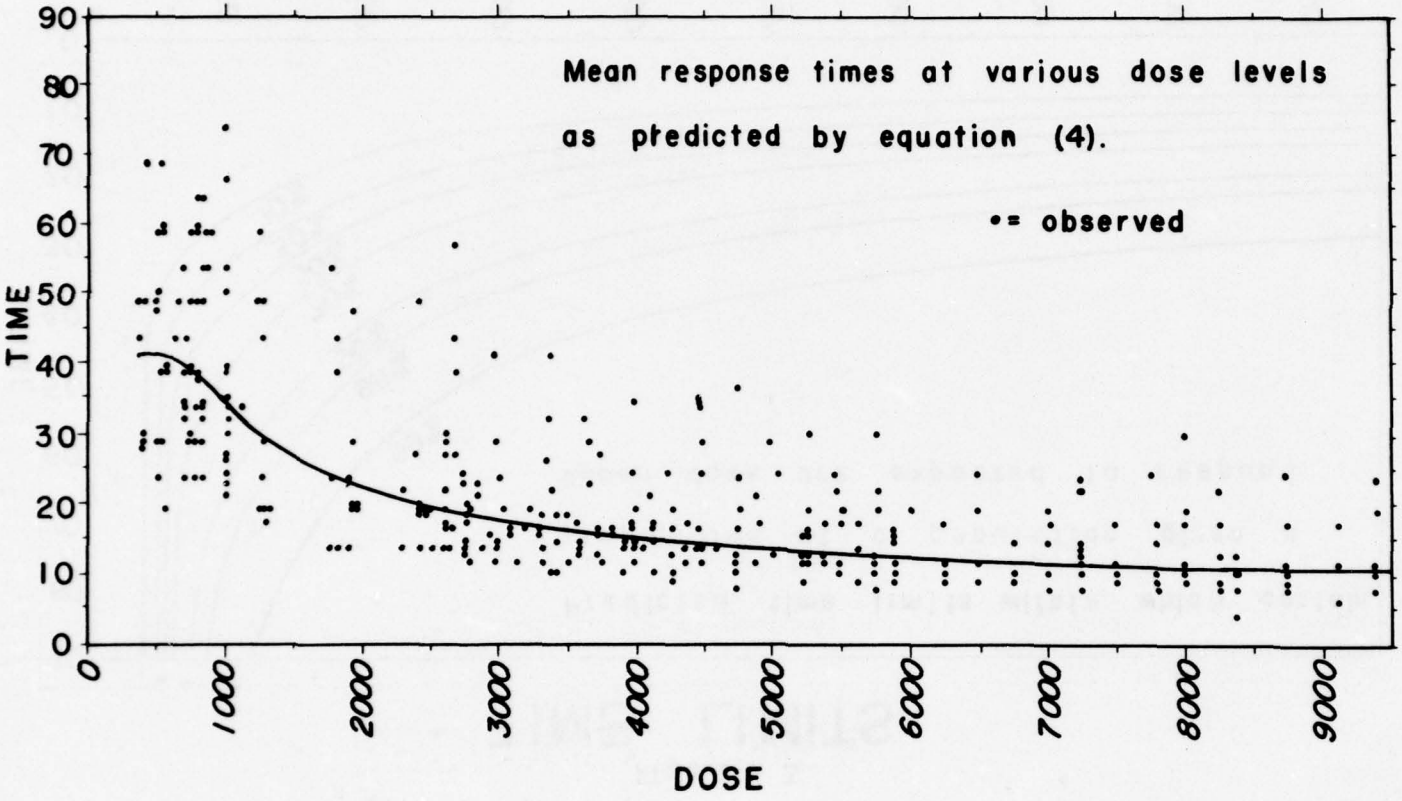
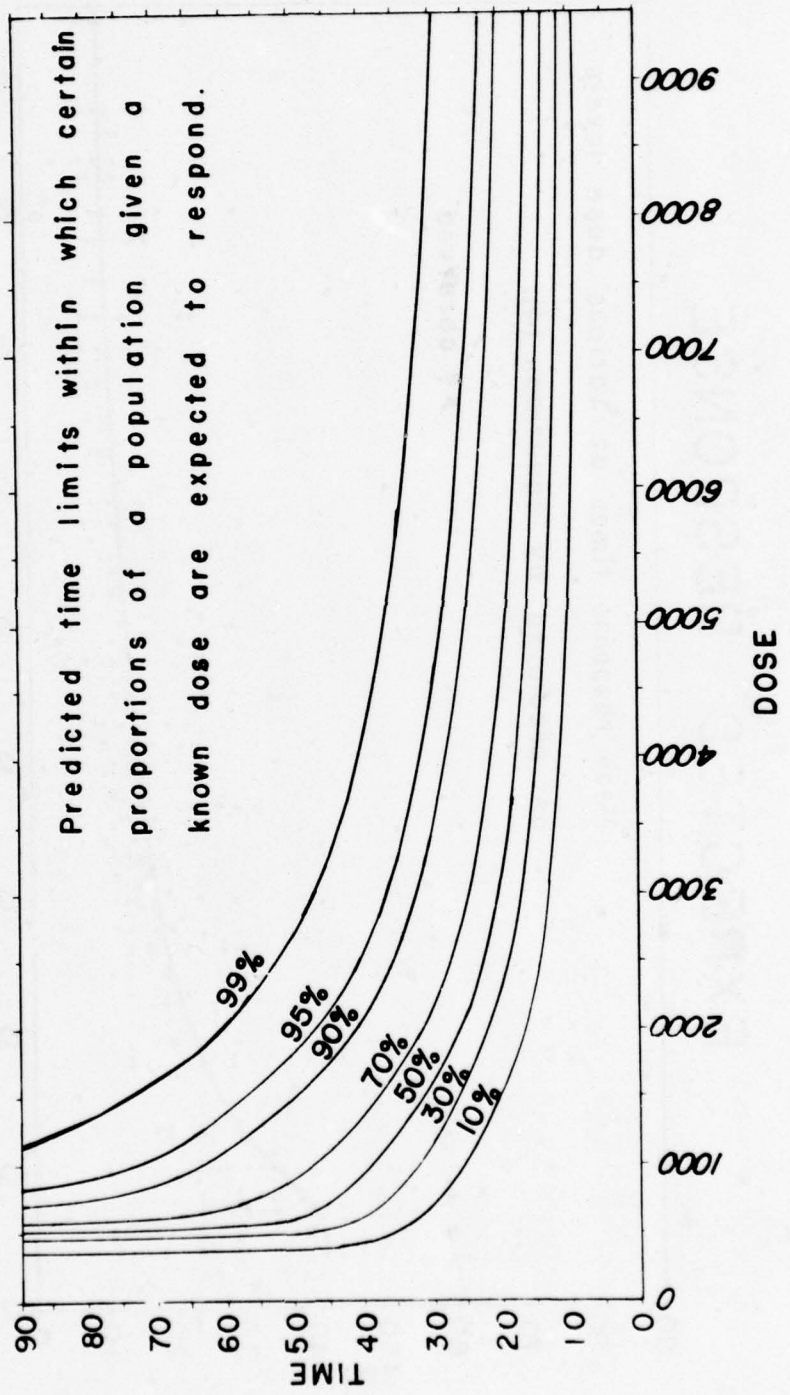


Figure 3

TIME LIMITS



LINEAR STRUCTURAL RELATIONSHIPS UNDERLYING THE DECOMPOSITION OF LEVINSTEIN H

By Henry Ellner

As the title of this paper indicates, we are going to be concerned with statistical and chemical relationships. But unfortunately there are very few of us who are well versed with the principles of both fields so that we can readily comprehend the interplay of ideas originating from the separate disciplines. If you can bear with me, either as statisticians or as chemists, you may find that actually chemical reactions can be expressed in terms of statistical regression equations and vice versa.

The analogous relationships between reaction and regression equations contained in Exhibit I on page 51, show chemical reactions depicting the decomposition of Levinstein process mustard gas and their statistical counterparts. The regression equations were developed first; the chemical equations followed from painstaking efforts to explain the structural information behind the regression coefficients. By structure, I mean the functional or physical connection behind the variates represented by regression equations. These equations are not limited to simple linear or multiple regression equations. At the end of this talk, you will find a bibliography pertaining to linear relationships involving instrumental variates and variables subject to error. These relationships estimate the true connections between the interacting variables, but as we shall see, the usual regression analyses will not mislead us.

Since we are going to depend largely on reference tables to make this talk clear, it would be well for us to become familiar with their contents. In Table I - Parts A and B, we have the analytical observations obtained on 52 samples of Levinstein H. They include the iron content of the sample as ferrous chloride, the purity of the sample as obtained by the distillation method, the freezing point of the sample and the acid concentration as percent hydrochloric acid. With the exception of the freezing point method, all tests were performed in accordance with the specification standard for Levinstein H. The freezing point method is a rapid test for the purity of Levinstein H, but not as precise or accurate as the distillation method. The observations in Part A are half

of the total, and are associated with the lowest 26 sample ranked by their iron content. The remaining half of the observations are in Part B associated with the upper 26 samples ranked by their iron content.

Table I - Part A

Experimental Results of Specification Tests on Levinstein H Stored in Steel Containers

Iron (%$FeCl_2$) (f)	Purity (%H) (h)	Freezing Point (^oC) (z)	Acid (%HCl) (a)
0.38	69.09	3.56	0.60
0.42	67.77	3.52	0.79
0.42	67.31	3.87	1.43
0.45	67.42	3.63	1.29
0.46	69.12	3.97	0.75
0.47	71.76	4.68	0.41
0.51	66.48	3.29	1.40
0.51	68.89	3.80	0.83
0.55	68.69	3.78	0.96
0.58	65.48	3.37	1.34
0.59	68.86	3.56	1.13
0.66	69.78	4.18	0.93
0.81	64.73	3.12	1.48
0.85	71.33	4.85	0.94
0.90	64.28	2.25	1.57
0.90	71.19	4.98	1.19
0.94	70.22	4.14	0.70
0.96	70.02	5.10	1.22
1.09	69.89	4.07	0.90
1.09	67.96	4.18	1.51
1.12	70.08	4.44	1.24
1.22	64.48	3.15	0.15
1.23	67.10	4.40	1.47
1.25	63.40	2.70	1.84
1.29	65.98	3.55	1.52
1.38	64.48	3.52	1.29
$\bar{x}_1(f) = 0.8088$	$\bar{x}_1(h) = 67.9150$	$\bar{x}_1(z) = 3.8331$	$\bar{x}_1(a) = 1.1108$

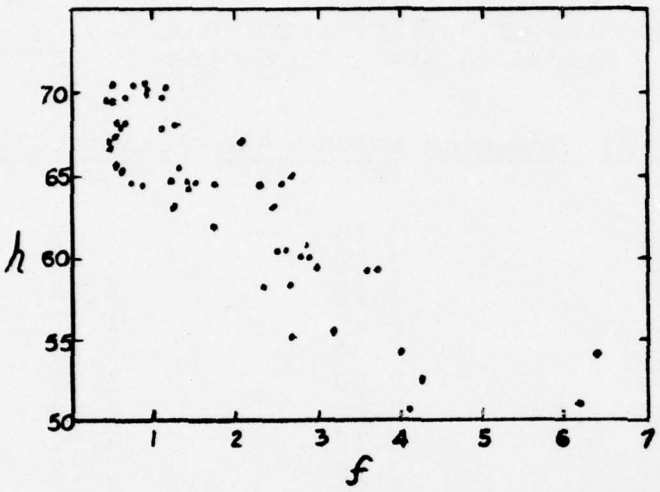
Table I - Part B

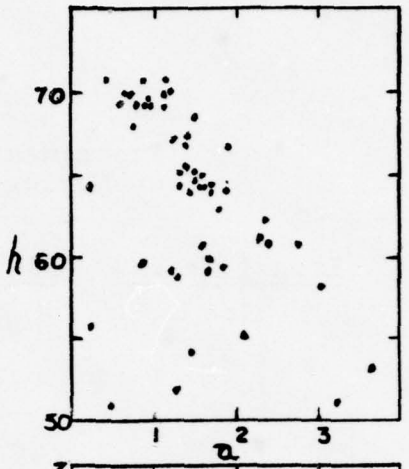
Experimental Results of Specification Tests on Levinstein H Stored in Steel Containers

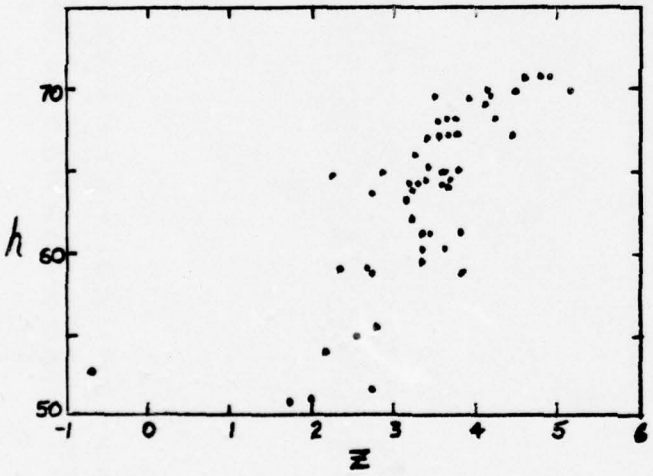
Iron ($FeCl_2$) (f)	Purity (H) (h)	Freezing Point (^oC) (z)	Acid (HCl) (a)
1.41	64.62	3.10	1.49
1.48	64.84	3.55	1.49
1.85	62.22	3.20	2.24
1.89	64.51	3.24	1.97
2.04	67.65	3.30	1.93
2.30	64.74	2.90	1.31
2.39	58.44	3.80	1.20
2.53	63.34	3.10	2.32
2.55	61.02	3.30	2.81
2.56	64.22	3.67	1.62
2.62	61.45	3.30	2.36
2.73	58.28	2.87	3.01
2.76	65.37	3.75	1.48
2.80	55.04	2.63	2.08
2.86	60.57	3.35	0.96
2.88	61.88	3.80	1.52
2.99	60.76	3.60	1.64
3.02	59.48	2.36	1.69
3.04	56.43	2.80	0.17
3.44	59.68	2.70	1.26
3.57	59.62	3.27	1.80
4.00	54.80	2.15	1.49
4.07	51.67	1.72	0.57
4.29	52.12	2.85	1.36
6.28	51.29	2.00	3.24
6.40	53.90	-0.85	3.62
$\bar{x}_2(f)=3.0288$	$\bar{x}_2(h)=59.9208$	$\bar{x}_2(z)=2.9023$	$\bar{x}_2(a)=1.7935$

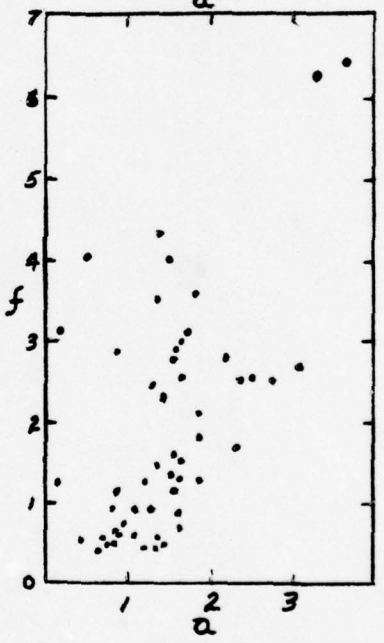
Next, we find correlation diagrams which show plots of the data of Table I. The upper four diagrams illustrate the scatter of values for pairs of variables. The symbols used are identified in the column headings of Table I. The abscissas of the two lowest correlation diagrams denote the pure mustard gas content of samples as predicted from multiple correlation equations. The associated observed mustard content of the samples is plotted in accordance with scale of the ordinate. The scatter of pairs of values, the

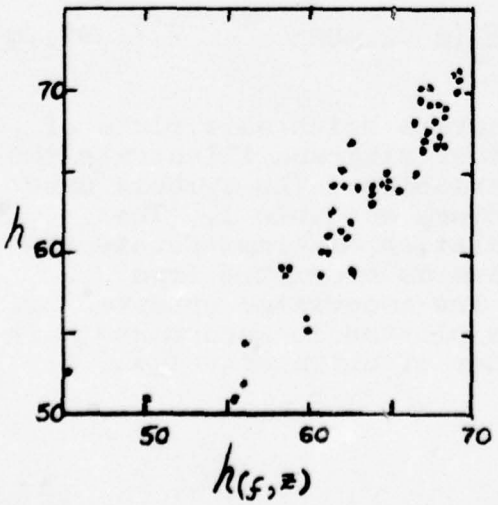
Correlation Diagrams

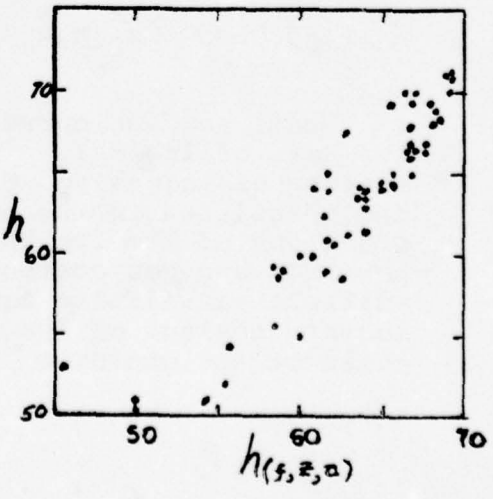












predicted and the observed, reveals the degree of lack of agreement.

Table II

Correlation and Partial Correlation Coefficients for Variables Shown in Table I

$r_{hf}=-0.880^{***}$	$r_{hf.z}=-0.743^{***}$	$r_{hf.a}=-0.847^{***}$	$r_{hf.za}=-0.749^{***}$
$r_{hz}=+0.760^{***}$	$r_{hz.f}=+0.4003^{**}$	$r_{hz.a}=+0.688^{***}$	$r_{hz.fa}=+0.428^{**}$
$r_{ha}=-0.459^{***}$	$r_{ha.z}=-0.128$	$r_{ha.f}=+0.1022$	$r_{ha.zf}=+0.1951$
$r_{fa}=+0.567^{***}$	$r_{fa.z}=+0.338^{*}$	$r_{fa.h}=+0.386^{**}$	$r_{fa.zh}=+0.365^{**}$
$r_{fz}=-0.712^{***}$	$r_{fz.h}=-0.140$	$r_{fz.a}=-0.597^{***}$	$r_{fz.ha}=-0.0360$
$r_{az}=-0.510^{***}$	$r_{az.h}=-0.279^{*}$	$r_{az.f}=-0.184$	$r_{az.hf}=-0.246$

* Significant at least on the 5 percent level
** Significant at least on the 1 percent level
*** Significant at least on the 0.1 percent level

Table II presents the total correlation and partial correlation coefficients for the four variables of Table I. The first column on the left contains the total correlation coefficients, each showing the degree of association on a scale ranging from zero to one between two variables, if the negative sign is ignored. Then looking at the adjacent columns from left to right we notice the effect on the correlation when the variables shown after the point or stop are fixed or eliminated statistically from interfering with the relation of the two variables concerned. The coefficients in columns 2, 3 and 4 are known as partial correlation coefficients.

Proceeding to Table III we find the simple linear and multiple regression equations which are going to be the basis for this talk. The last two equations were used to predict the pure mustard content of Levinstein H samples shown in the lowest two correlation diagrams. The equation depicting h as the dependent variable and x_f and x_z as the independent variables is portrayed on the lower left and the equation depicting h as dependent upon x_f, x_z and x_a

is portrayed on the lower right of the correlation diagrams.

The multiple correlation coefficients which are defined by analogy with the simple correlation coefficient are given adjacent to the pertinent multiple regression equation. They are used as a measure of the strength of the association of h with x's.

Table III

Simple Linear and Multiple Regression Equations and Coefficients for Variables of Table I

$h = 70.36-3.36\ x_f,$	$r = -0.880$
$*h = 71.13-3.88\ x_f,$	$r = -0.875$
$h = 69.94-3.48\ x_f+0.46\ x_a,$	$R_{h(fa)} = 0.881$
$h = 63.67-2.62\ x_f+1.57\ x_z,$	$R_{h(fz)} = 0.900$
$h = 62.38-2.79\ x_f+1.70\ x_z+0.81\ x_a$	$R_{h\,(fza)} = 0.904$

*Observations associated with and including the two highest ranking iron content determinations were eliminated.

Table IV

Sums of Squares and Sums of Products of Deviations of Variables from Their Means

From data of Table I:

$Sx_h'^2 = 1506.5313$	$Sx_f'^2 = 103.4699$
$Sx_z'^2 = 44.9781$	$Sx_a'^2 = 25.2869$
$Sx_h'x_f' = -347.7282$	$Sx_h'x_z' = +197.8965$
$Sx_h'x_a' = -89.6539$	$Sx_f'x_z' = -48.5891$
$Sx_f'x_a' = +28.9888$	$Sx_z'x_a' = -17.2160$

Analysis of Variance Table V

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Squares
Variation explained by x_f	1168.61	1	1168.61
Increment explained by addition of x_a	1.15	1	1.15
Total explained by x_a and x_f together	1169.76		
Residual	336.77	50	6.74
Total	1506.53		

Analysis of Variance Table VI

Source of Variation	Sums of Squares	Degrees of Freedom	Mean Squares
Variation explained by x_f	1168.61	1	1168.61
Increment explained by addition of x_z	53.40	1	53.40
Total explained by x_f and x_z together	1222.01		
Residual	284.52	50	5.69
Total	1506.53		

Analysis of Variance and Covariance Table VII
Within and Between Groups by Wald's Method

Variance of x(f)	S. s.	D. f.	M. s.
(IV) Within groups	$Ns'_x = 39.40$	N-2 = 50	7.88
(I) Between groups	$(N/4)(\bar{x}_2-\bar{x}_1)^2 = 64.07$	1	64.07
Total	$Ns^2_x = 103.47$	N-1 = 51	20.29

Variance of y(h)	S. s.	D. f.	M. s.
(VI) Within groups	$Ns'_y = 675.73$	N-2 = 50	13.51
(III) Between groups	$(N/4)(\bar{y}_2-\bar{y}_1) = 830.81$	1	830.81
Total	$Ns_y = 1506.54$	N-1 = 51	29.54

Covariance	S. p.	D. f.	M. p.
(V) Within groups	$Ns'_{xy} = 117.01$	N-2 = 50	-2.34
(II) Between groups	$(N/4)(\bar{x}_2-\bar{x}_1)(\bar{y}_2-\bar{y}_1) = -230.72$	1	-230.72
Total	$Ns_{xy} = -347.72$	N-1 = 51	6.82

$$b = \frac{-347.73}{103.47} = -3.36$$

$$b' = \frac{-347.73}{1506.54} = -2.31 \qquad 1/b' = -4.33$$

$$\hat{\beta} = \frac{-230.72}{64.07} = -3.60$$

$$\hat{\sigma}_\varepsilon^2 = (39.40 - \frac{117.01}{3.60})/50 = 0.13814$$

$$\hat{\sigma}_\varepsilon = 0.372 \text{ (error in } x_f\text{)}$$

$$\hat{\sigma}_\eta^2 = 675.73 - (117.01)(3.60)/50 = 5.0899$$

$$\hat{\sigma}_\eta = 2.256 \text{ (error in } y_h\text{)}$$

Table IV provides the statistician with derived data for calculating the regression coefficients of the simple linear and multiple regression equations of Table III. These data also serve as the basis for the two Analysis of Variance Tables V and VI. Statisticians will comprehend the use of the tables and the chemists will do best to skip to Exhibit I on page showing the analogous relationships between reaction and regression equations.

I am not so sure that all of the statisticians will be familiar with the analysis of variance and covariance Table VII so that I shall note in passing that they consult E. S. Keeping's paper listed in the selected references.

Exhibit I contains the real meat of this paper. Here we have postulated chemical reactions and stoichiometric relations suggested by the regression equations. From organic and physical chemical principles and research reported in the literature, the reactions postulated seem likely.

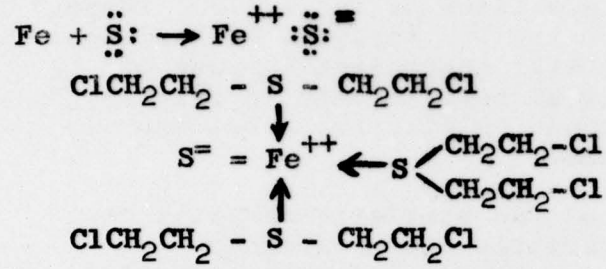
Now I believe we are ready to delve into the stoichiometric relationships and structural relationships underlying the deterioration of Levinstein H under normal storage. Levinstein H is not a pure compound. It is a very complex mixture containing from 68 to 73 percent pure mustard and the balance impurities consisting of polysulfides which degredate in time to precipitate sulfur and form rearrangements of the sulfur atom linkages.

Pure mustard is represented by the formula $ClCH_2CH_2$-S-$ClCH_2CH_2$. Replacement of the single sulfur atom by two or more sulfur linkages results in a polysulfide. The polysulfide believed to be generally present in the highest concentration in newly made Levinstein H is represented by the formula: $(ClCH_2CH_2)_2S_7$.

The nature of polysulfides has been ably discussed in a published paper by Dr. Macy of the Chemical Corps. This paper stated that the remarkably rapid deterioration of Levinstein H in small steel containers under tropical conditions (100-150° F) is thought to be due to the oxidizing nature of the sulfur in the polysulfides upon the iron content of the container. This reaction sets off a chain of events which culminates in a substantial loss of pure mustard with the concomittant formation of a tarry material even more vesicant than the original.

Exhibit I

Analogous Relationships Between Reaction and Regression Equations



(1)

(2)

Simple linear regression equation (2 variables)

$$h = 70.36 - 3.36x_f \qquad (2a.)$$

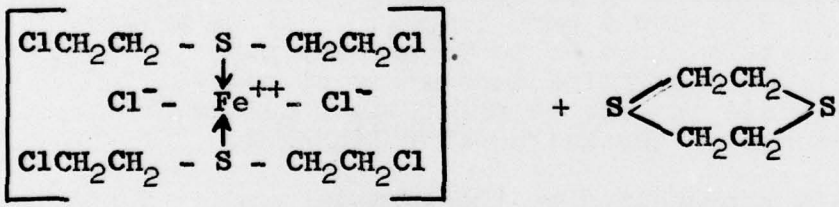
Wald's structural relation (2 variables)

$$h = 70.83 - 3.60x_f \qquad (2b.)$$

Stoichiometric equivalent:

$$3(CH_2CH_2Cl)_2S \rightleftharpoons FeCl_2 \qquad (2c.)$$

3.76% change in pure H $\rightleftharpoons$ 1% change in $FeCl_2$.



(3)

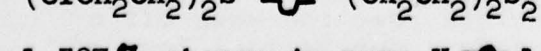
Multiple linear regression (3 variables)

$$h = 63.67 - 2.62x_f + 1.57x_z \qquad (3a.)$$

Stoichiometric equivalents:

$$2(CH_2CH_2Cl)_2S \rightleftharpoons FeCl_2 \qquad (3b.)$$

2.51% change in pure H $\rightleftharpoons$ 1% change in $FeCl_2$.



(3c.)

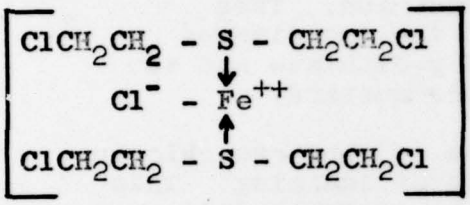
1.327% change in pure H $\rightleftharpoons$ 1% change in p-dithiane.

From Raolt's Law $\Delta t = 36.4\ N$ for pure H

$\therefore$ 1°C change in $\Delta t \rightleftharpoons$ 0.0275% change in N

and 1.585% change in pure H $\rightleftharpoons$ 1°C change in Δt,

where $\Delta t = \Delta \mathcal{Z}$, and N is mole fraction of impurity.



$$Cl^- + S\begin{matrix}\diagup CH_2CH_2 \diagdown \\ \diagdown CH_2CH_2 \diagup\end{matrix}S \qquad (4)$$

Multiple linear regression (4 variables)

$$h = 62.38 - 2.79x_f + 1.70x_{\mathcal{Z}} + 0.81x_a \qquad (4a)$$

Stoichiometric equivalents:

$$2.223(CH_2CH_2Cl)_2S \rightleftharpoons FeCl_2, \text{ since } \frac{M.W.(Cl)}{M.W.(CH_2CH_2Cl)_2S} = 0.223 \qquad (4b)$$

2.79% change in pure H $\rightleftharpoons$ 1% change in $FeCl_2$

$$(1.223)0.777(CH_2CH_2Cl)_2S \rightleftharpoons (CH_2CH_2)_2S_2 \qquad (4c)$$

1.258% change in pure H $\rightleftharpoons$ 1% change in p-dithiane.

From Raolt's Law, $\Delta t = 36.4\ N$ for pure H,

where $\Delta t = \Delta \mathcal{Z}$, and N is the mole fraction of impurity.

$\therefore$ 1.67% change in pure H $\rightleftharpoons$ 1°C change in $\mathcal{Z}$

$$(0.223)(0.777)(CH_2CH_2Cl)_2S \rightleftharpoons HCL \qquad (4d)$$

.755% change in pure H $\rightleftharpoons$ 1% change in HCL

If we now refer to Exhibit I we can follow through with the postulated series of reactions. Equation (1) shows the chemical equation leading to the ferrous ion build-up in a Levinstein H sample. The next equation assumes that 3 moles of pure mustard react with the ferrous ions to obtain an intermediate coordinated compound, showing dative bonds between three sulfur atoms and the ferrous ion. Then equation (3) shows the decomposition of the coordinated structure to form the chemical compound p-dithiane and the obelated structure with two moles of pure mustard.

Equation (4) is similar to equation (3) but one chlorine atom is shown outside of bracket capable of ionizing. This reaction as well as the preceding ones is somewhat different from the mechanisms proposed by chemists who worked with pure reagents. Bell and his co-workers postulated the formation of ethylene chloride as well as p-dithiane, and Fuson's group at the University of Illinois substantiated this finding. But both research groups used pure mustard in glass vessals which were heated at elevated temperatures of 150°C to 180°C from 18 to 48 hours. They then inferred that a similar reaction occurred at the so-called "ordinary" temperature. Fuson also arrived at a mechanism for the formation of the tarry decomposition product of Levinstein H. Again his group heated pure mustard and found condensation products which Fuson thought would account for some of the highly insoluble material present in samples of heated mustard gas.

Apparently then the postulated series of equations must be substantiated if they are to be accepted. The proof, if it may be considered such, lies in the data of Table I, Parts A and B. These data represent a total stock of some 850,000 pounds of Levinstein H, accumulated as a result of World War II, which were transferred to steel one-ton containers for convenience of storage. Nothing is known of their previous history except that at one time the batches of mustard met the specification quality standards.

To obtain a record of the current quality, a total of 52 samples were selected at random and subjected to the specification tests as well as to a freezing point test used for periodic surveillance of the stock. The data are shown ranked in accordance with the ferrous chloride concentration of each sample. This order was found useful in performing subsequent statistical analyses, and can now serve to orient non-statisticians in a casual review of the data.

The correlation diagrams illustrate the connection between the variables of Table I. The symbols used to denote the variables are supposed to be mnemonic with h for pure mustard, a for acidity, z for freezing point of the mustard and f for the ferrous chloride concentration. But if one is not careful, one may think that f represents the freezing point. The letter i is reserved by statisticians for denoting an individual value so that its use as a symbol for iron would be confusing to them.

The scatter of points may confound the chemist, but it is a perfect delight to the statistician who can now ply his trade. And we can see how foreboding the charts are when we look at Table II. At least the statisticians will be struck by the highly significant correlations which are rendered insignificant when the secondary variables shown after the point or stop are fixed. For the benefit of the chemists as well as the statisticians, I should state that these symbols are in accordance with the well-known text of Yule and Kendall.

What do the correlation coefficients tell us? For one thing, when the ferrous ion concentration is held constant, we can't predict the mustard content from knowledge of the acid concentration. Similarly, the ferrous ion concentration does not affect the depression of the freezing point of Levinstein H when the pure mustard content and the acid concentration are held constant. These are weird bits of knowledge, but they fit into the mosaic representing the pattern postulated by the series of chemical reactions appearing in the reference sheets. All of the correlation coefficients derived from the data of Table I have meaning to the experienced observer.

The technical importance of the associated variables is shown more absolutely by the appropriate regression coefficients of the simple and multiple regression equations shown in Table III. The regression coefficients provide the average rate of change of the dependent variable with a unit change in the independent variable or variables. If the reality of these coefficients can be established, we are then in a strong position to discuss their relation with the postulated chemical reactions.

The existence of the regression coefficients can best be determined by testing for their significance in an analysis of variance table. The two analysis of variances tables, V and VI, show that there can be no doubt of the effect of the ferrous chloride concentration upon the decomposition of the pure mustard content of Levinstein H. However, the acidity of the sample does not appear to be of any consequence. With respect to the freezing point, it appears to be a fair indicator of the purity of the sample.

When the prediction equation for the pure H content of Levinstein mustard involved four variables, another approach which can be comprehended by the chemist is to set limits of uncertainty, that is, 95 percent confidence limits for the regression coefficients. We find by a modification of Fisher's that these limits for the coefficients of the equation, $h = 62.38 - 2.79\ x_f + 1.70\ x_z + 0.81\ x_a$ are the following: -2.79 ± 0.71, $+1.70 \pm 1.04$ and $+0.81 \pm 1.18$. The last coefficient includes the value of zero and is therefore deemed nonsignificant. Again the ferrous chloride concentration and the freezing point determination are interrelated with the decomposition of Levinstein H with respect to its purity.

Since it cannot be presumed that the ferrous ions react completely with the pure mustard contained in Levinstein H, we need another statistical method to give the best quantitative estimate of the functional relationship between the two reactants. We shall use the simple and useful method of Wald's, although statisticians may want to apply the more advanced procedures cited in selected references included in the reference sheets. Wald's method has recently been given a thorough review by Kiefer and Wolfowitz, but Keeping has provided a practical approach that checks the assumptions underlying Wald's method. This procedure is set out in the Analysis of Variance and Covariance Table VII. The analysis yields the sought-for relation between the pure mustard content of Levinstein H and the ferrous ions expressed as ferrous chloride. A one percent increase in the ferrous chloride concentration effects a 3.60 percent decrease in the purity of Levinstein H. We also have estimates of the errors in the ferrous ion determination and the analytical procedure for determining the purity of mustard. These errors, which include the non-reactive portions of the variables, have been stripped from observed values to provide a valid functional relationship between the reactants.

Now that we have a good quantitative estimate of functional relationship between pure mustard and the ferrous ions, we are ready to tackle the analogous relationships between the reaction

and regression equations. Let us examine the reaction equations and the regression equations side by side.

The stoichiometric equivalent given by (2a) is derived from equation (2). The 3.76 percent change in pure mustard, which is equivalent to a 1 percent change in ferrous chloride concentration, is analogous to regression coefficients of equations (2a) and (2b).

Now the chemical reaction depicted by (3) is similarly analogous to regression equation (3a), but we have two coefficients to consider. From the stoichiometric equivalent (3b) we find that the change in pure mustard equivalent to a 1 percent change in ferrous chloride concentration is close to the coefficient 2.62. Since p-dithiane was not determined analytically, its effect is observed by its presence as an impurity added to the pure mustard content of Levinstein H. The freezing point is thereby affected which, in turn, indicates a drop in the mustard content. It all resembles a Rube Goldberg cartoon - but the entire train of events can be followed through the stoichiometric equivalent (3c) and the application of Raoult's Law. The analogous relationship between the regression coefficient of 1.57 and the change in pure mustard corresponding to a one degree change in the freezing point is quite apparent.

The reaction (4) is analogous to the regression equation (4a) which involves four variables. To demonstrate this relationship we recall that (3) showed that 2 moles of pure mustard combined with one mole of ferrous chloride, $FeCl_2$, and that one mole of mustard yielded one mole of p-dithiane, accounting for the decomposition of three moles of pure mustard. The problem now is to repartition the three moles of mustard to account for the magnitude of each of the three regression coefficients confronting us in regression equation (4a).

Considering that portion of reaction (4) within the square brackets we obtain the stoichiometric equivalent (4b), which states that 2.223 moles of mustard is equivalent to one mole of ferrous chloride. The equivalency reflects that one mole of ferrous chloride, $FeCl_2$, less one mole of chloride ion combines with two moles of mustard. From this relationship we readily find that a 2.79 percent decrease in pure H concentration is manifested by a 1 percent increase in the ferrous ion concentration reported as ferrous chloride.

As we have accounted for 2.223 moles of mustard, there

remains 0.777 moles to be partitioned between the chloride ion shown outside of the square brackets of reaction (4), and the p-dithiane. The presence of the chloride ion is indicated by hydrolysis when acidity of the Levinstein H sample is determined, and by depression of the sample's freezing point. The only effect of p-dithiane is to lower the freezing point of the Levinstein H. Now since a total of 3 moles of mustard must be accounted for as a result of reaction (2) the following partitioning satisfies the combined effects:

$$2.223 + \left[(1.223)\ (0.777)\right] - \left[(0.223)\ (0.777)\right] = 3.000.$$

The quantity within the square brackets on the far left represents the moles of mustard yielding p-dithiane as shown by (4c). The quantity within the square brackets preceded by a negative sign represents the moles of mustard yielding hydrochloric acid as shown by (4d).

The derived stoichiometric equivalents from the above partitioning of mustard undergoing decomposition agree remarkably well with the corresponding partial regression coefficients as to magnitude and sign. Utilizing Raoult's Law we derive the relation that a 1.67 percent decrease in pure mustard is equivalent to a one degree Centigrade decrease in the freezing point of the crude mustard. We also obtain the relation that a 0.76 percent increase in pure mustard corresponds to 1.0 percent increase in hydrochloric acid. The latter seems contradictory to observed effects but reflection will disclose that the apparent increase in purity due to HCl serves to counteract the excessive decrease in purity represented by coefficients associated with the iron and freezing point variables of equation (4a). When these coefficients are compared with corresponding coefficients of multiple regression equation (3a) it will be noted that the latter are substantially smaller.

Unravelling the mechanism underlying the decomposition of Levinstein H is an illustration of the general approach in applying regression, functional and structural relationships to discover scientific laws. In the field of chemistry, linear structural relationships are of fundamental importance in that chemical elements are natural building blocks which can be grouped to form linear structures. Moreover the properties of molecular compounds are usually proportional to their concentration, especially under certain restrictions. Consequently, the analysis of apparently complex relationships can be handled quite simply by applying statistical

techniques developed for the special purpose of determining structural or functional relationships. In the special case of a decomposition reaction the intermediate stages can be revealed by noting the analogous stoichiometric equivalents and the corresponding regression coefficients of successive terms of simple and multiple regression equations.

Selected References on Determining Structural or Functional Relationships

A. Wald, "The fitting of straight lines if both variables are subject to error", Ann. Math. Stat., Vol. 11 (1940), pp. 284-300.

D. V. Lindley, "Regression lines and the linear functional relationship", Suppl. J. Roy, Stat. Soc., Vol. 9 (1947), pp. 218-244.

R. C. Geary, "Determination of linear relations between systematic parts of variables with errors of observation the variances of which are unknown", Econometrica, Vol. 17 (1949), pp. 30-58.

M. S. Bartlett, "Fitting a straight line when both variables are subject to error", Biometrics, Vol. 5 (1949), pp. 207-212.

Olav Reirsøl, "Identifiability of a linear relation between variables which are subject to error", Econometrica, Vol. 18 (1950) pp. 375-389.

J. W. Tukey, "Components in Regression", Biometrics, Vol. 7 (1951), pp. 33-69.

M. G. Kendall, "Regression, structure and functional relationship", Part I, Biometrika, Vol. 38 (1951), pp. 11-25.

J. Neyman and E. L. Scott, "On certain methods of estimating linear structural relation between two variables", Ann. Math. Stat., Vol. 22 (1951), pp. 352-361.

M. G. Kendall, "Regression, structure and functional relationship", Part II, Biometrika, Vol. 39 (1952), pp. 96-108.

D. V. Lindley, "Estimation of a functional relationship", Biometrika, Vol. 40 (1953), pp. 47-49.

C. A. Bennett and N. L. Franklin, Statistical Analysis in Chemistry and the Chemical Industry, John Wiley & Sons, Inc., New York (1954), pp. 465-469.

J. Wolfowitz, "Estimation of the components of stochastic structures", Proc. Nat. Acad. Sci., U.S.A., Vol. 40, No. 7 (1954), pp. 602-606.

E. S. Keeping, "Note on Wald's method of fitting a straight line when both variables are subject to error", Biometrics, Vol. 12 (1956), pp. 445-448.

J. Kiefer and J. Wolfowitz, "Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters", Ann. Math. Stat., Vol. 27 (1956), pp. 887-906.

Selected References on Levinstein Mustard Gas

E. V. Bell, G. M. Bennett, and A. L. Hock, J. Chem. Soc., Vol. 131 (1927), p. 1803.

M. Sartori, The War Gases, D. Van Nostrand, New York, (1939).

R. C. Fuson, et al, J. Org. Chem., Vol. 11 (1946), pp. 469-517.

R. Macy, et al, "The Polysulfides in Levinstein Process Mustard Gas", Science, Vol. 106 (1947), pp. 355-359.

Proceedings of the
Third Annual Statistical
Engineering Symposium

Submitted by:

DAVID R. HOWES
Symposium Secretary

T. M. VINING
Symposium Chairman

Approved By:

LOUIS E. GARONO
Deputy for Engineering

WILLIAM J. ALLEN, JR.
Colonel, CmlC
Commanding Officer

U. S. Army Chemical Corps Engineering Command
Army Chemical Center, Maryland

DISTRIBUTION

1. U. S. Army Chemical Corps Engineering Command distribution list for Statistical Engineering Publications (1 copy each)
2. Symposium Attendance List (1 copy each)
3. Office of Deputy Chief Chemical Officer for Scientific Activities (10 copies)
4. Director of Research and Development, U. S. Army (4 copies)
5. Armed Services Technical Information Agency (5 copies)