AD612886

CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION, CFSTI
INPUT SECTION 410.11

LIMITATIONS IN REPRODUCTION QUALITY OF TECHNICAL ABSTRACT BULLETIN DOCUMENTS, DEFENSE DOCUMENTATION CENTER (DDC)

☐ 1. AVAILABLE ONLY FOR REFERENCE USE AT DDC FIELD SERVICES. COPY IS NOT AVAILABLE FOR PUBLIC SALE.

☒ 2. AVAILABLE COPY WILL NOT PERMIT FULLY LEGIBLE REPRODUCTION. REPRODUCTION WILL BE MADE IF REQUESTED BY USERS OF DDC.

☒ A. COPY IS AVAILABLE FOR PUBLIC SALE.

☐ B. COPY IS NOT AVAILABLE FOR PUBLIC SALE.

☐ 3. LIMITED NUMBER OF COPIES CONTAINING COLOR OTHER THAN BLACK AND WHITE ARE AVAILABLE UNTIL STOCK IS EXHAUSTED. REPRODUCTIONS WILL BE MADE IN BLACK AND WHITE ONLY.

TSL-121-2/64

DATE PROCESSED: 4-2-65

PROCESSOR: B. Lee

STUDIES IN RESEARCH METHODOLOGY

VI. THE CENTRAL LIMIT EFFECT FOR A VARIETY OF POPULATIONS AND THE ROBUSTNESS OF Z, t, AND F

JAMES V. BRADLEY, PhD

FOREWORD

This report was prepared by James V. Bradley of the Presentation of Information Branch, Human Engineering Division, Behavioral Sciences Laboratory, working under Project No. 7184, "Human Performance in Advanced Systems," Task No. 718401, "Criteria for the Design and Arrangement of Displays." This report is the sixth in a series by the same author, investigating the applicability of "standard," i.e., general, research methods to the idiosyncratic conditions characterizing experimentation in the behavioral sciences. The previous reports of the series are listed in the bibliography.

The report is a sequel to the author's Ph.D. thesis study and as such is the once-removed beneficiary of extremely helpful suggestions from Dr. Benjamin J. Winer of the Psychology and Statistics Departments and Dr. James A. Norton of the Statistics Department, Purdue University.

Sampling was performed at the Digital Computation Division, Directorate of Systems Dynamic Analysis, using an IBM 7090 computer. The problem was programmed by Mr. Tom Duvall. Mathematical analysis relevant to the generation of the F tables used was supplied by Mr. Edwin Godfrey, and methods of generating pseudo-random numbers were supplied and tested by Mr. Clem Grabner and Mr. Edwin Godfrey. An ancillary problem yielding the data points for Figures 65-70 was programmed by Mr. John Smith. The efficiency, cooperativeness, and professional talent of the aforenamed contributed greatly to the success of the study.

Published tables do not contain all of the t values needed for this study. The required values were generously supplied by Dr. Donald B. Owen of the Sandia Corporation who used an existing computer program to obtain the t values unobtainable from published tables because of the high degrees of freedom involved. The writer is greatly indebted to Dr. Owen for the highly accurate values supplied.

This technical report has been reviewed and is approved.

WALTER F. GRETHER, PhD
Technical Director
Behavioral Sciences Laboratory

ABSTRACT

The robustness of Z, t and F tests was studied by obtaining 1220 empirical sampling distributions (some of which were later combined) each consisting of 10,000 values of the test statistic obtained under a unique combination of sampling conditions. Conditions investigated, both alone and in combination, were: population shapes (nonnormal, normal, or some nonnormal and others normal), population variances (all σ^2, or some σ^2 and others $\sigma^2/4$), relative sample sizes (for two-sample tests N,N; 2N,N; 3N,N; or 2N,2N; for other multi-sample tests N,N,N; 2N,N,N; or N,N,N,N), and absolute sample size (N assumed values of 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, and in four special cases 2048 and 4096). Robustness at nominal significance levels of .05, .01, and .001 was examined for both left-, right- and two-tail tests. The Central Limit effect upon means of samples from populations differing considerably in shape and "degree of nonnormality" was revealed in the robustness of certain Z statistics. Results showed that robustness varies drastically with conditions which are mentioned neither in the statement of the assumption which was violated nor in the claim, as usually propounded, that the test is robust against that violation. In general, results were iconoclastic with regard to general, unqualified, allegations of robustness.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
POPULATIONS SAMPLED	2
STATISTICS INVESTIGATED	5
SAMPLING SCHEME	8
IDENTIFYING CODES FOR TEST STATISTICS	10
POPULATIONS AND STATISTICS RELEVANT TO THE CENTRAL LIMIT THEOREM	23
RESULTS	26
I. Manner of Presentation of Results	26
II. The Central Limit Effect as Influenced by Population Shape	28
III. Robustness of the One-Sample t Test	42
IV. Robustness of Multi-Sample Tests	42
DISCUSSION	80
I. The Central Limit Effect	80
II. Robustness of One-Sample t Tests	90
III. Robustness of Two-Sample Z Tests	91
IV. Robustness of Two-Sample t Tests	101
V. Robustness of Two-Sample F Tests for Equal Variances	110
VI. Robustness of the Analysis-of-Variance F Test	111
VII. Manner in which Individual Factors Influence Robustness	164
Nonnormality	166
Heterogeneity of Population Shape	166
Heterogeneity of Variance	166
Absolute Sample Size	167
Relative Sample Size	168
Number of Samples	169
Type of Test	169
Significance Level	170
Location of Rejection Region	170
VIII. General Comments	171
SUMMARY	173
CONCLUSIONS	195
BIBLIOGRAPHY	198
APPENDIX I Formulas Used to Calculate Test Statistics	200
APPENDIX II Sampling	204

TABLE OF CONTENTS (Contd)

Page

APPENDIX III
Checks. 212

APPENDIX IV
Photographed Tables of Original Output Data. 218

APPENDIX V
Tolerance Limits for Empirical Cumulative Probabilities. 273

APPENDIX VI
Influence of Population Moments upon the Robustness of the Sample Mean. 276

APPENDIX VII
Derivational Note. 287

LIST OF TABLES

TABLE I. 11

TABLE II 6

TABLE III. 22

TABLE IV 94

TABLE V. 180

TABLE VI 180

TABLE VII. 215

TABLE VIII 275

LIST OF ILLUSTRATIONS

Figure

1 The X and Y Populations 4

2 The X Population. 30

3 Central Limit Effect upon $\bar{X}$ 31

4 The X-X Difference-Score Population 32

5 Central Limit Effect upon $\overline{X-X}$ 33

6 The X-Y Difference-Score Population 34

LIST OF ILLUSTRATIONS (Contd)

Figure		Page
7	Central Limit Effect upon $\overline{X-Y}$	35
8	The X-A Difference-Score Population	36
9	Central Limit Effect upon $\overline{X-A}$	37
10	The X-B Difference-Score Population	38
11	Central Limit Effect upon $\overline{X-B}$	39
12	The Y-A Difference-Score Population	40
13	Central Limit Effect upon $\overline{Y-A}$	41
14	Robustness of TX.	43
15	Robustness of ZXX NN and TXX NN	47
16	Robustness of ZXX 2N2N and TXX 2N2N	48
17	Robustness of ZXX 2NN and TXX 2NN	49
18	Robustness of ZXX 3NN and TXX 3NN	50
19	Robustness of ZXY N3N and TXY N3N	51
20	Robustness of ZXY N2N and TXY N2N	52
21	Robustness of ZXY 2N2N and TXY 2N2N	53
22	Robustness of ZXY 2NN and TXY 2NN	54
23	Robustness of ZXY 3NN and TXY 3NN	55
24	Robustness of ZXA N3N and TXA N3N	56
25	Robustness of ZXA N2N and TXA N2N	57
26	Robustness of ZXA 2N2N and TXA 2N2N	58
27	Robustness of ZXA 2NN and TXA 2NN	59
28	Robustness of ZXA 3NN and TXA 3NN	60
29	Robustness of ZXB N3N and TXB N3N	61
30	Robustness of ZXB N2N and TXB N2N	62
31	Robustness of ZXB 2N2N and TXB 2N2N	63

LIST OF ILLUSTRATIONS (Contd)

Figure		Page
32	Robustness of ZXB 2NN and TXB 2NN	64
33	Robustness of ZXB 3NN and TXB 3NN	65
34	Robustness of ZYA N3N and TYA N3N	66
35	Robustness of ZYA N2N and TYA N2N	67
36	Robustness of ZYA 2N2N and TYA 2N2N	68
37	Robustness of ZYA 2NN and TYA 2NN	69
38	Robustness of ZYA 3NN and TYA 3NN	70
39	Robustness of TYB	71
40	Robustness of TD 2N and (repeated graphs) TXY 2N2N.	72
41	Robustness of RXX 2N2N and RXY 2N2N	73
42	Robustness of FXX (lower graphs are repeats).	74
43	Robustness of FXY	75
44	Robustness of FXA	76
45	Robustness of FXB	77
46	Robustness of FYA	78
47	Robustness of FYB	79
48	Influence of Population Asymmetry (about the Mean) upon the Central Limit Effect	85
49	Influence of Population Asymmetry (about the Median) upon the Central Limit Effect	86
50	Influence of Population Asymmetry (about the Mode) upon the Central Limit Effect	87
51	Dominance of Mean over Median or Mode as the Relevant Axis. .	88
52	Dominance of Asymmetry over "nonnormality" in Influencing Central Limit Effect.	89
53	Robustness of the Sample Variance	110
54	Diminishing Variability of Denominator, Relative to Numerator, of F as Sample Size Increases.	113

Figure		Page
55	Ultimate Robustness of F when Samples are of Equal Size and There are No More than Two Different Population Variances .	128
56	Same Data as Figure 55, Plotted to Different Abscissa Scale .	129
57	Critical Section of Figure 56, Plotted to Linear Ordinate Scale	130
58	Ultimate Robustness of F when Samples are of Equal Size and all Samples but One Come from Populations with the Same Variance	132
59	Same Data as Figure 58, Plotted to Different Abscissa Scale .	133
60	Critical Section of Figure 59, Plotted to Linear Ordinate Scale	134
61	Ultimate Robustness of t Statistic.	144
62	Same Data as Figure 61, Plotted to Different Abscissa Scale .	145
63	Critical Section of (unplotted first quadrant of) Figure 62, Plotted to Linear Ordinate Scale.	146
64	Relationship between $\frac{\text{Ult. Stand. Dev. (t)}}{\text{Ult Normal-Theory Stand. Dev. (t)}}$ and ρ/α	148
65	Ultimate Robustness of One-Tailed t Test Conducted at .05 Level of Significance	150
66	Ultimate Robustness of Two-Tailed t Test Conducted at .05 Level of Significance	151
67	Ultimate Robustness of One-Tailed t Test Conducted at .01 Level of Significance	152
68	Ultimate Robustness of Two-Tailed t Test Conducted at .01 Level of Significance	153
69	Ultimate Robustness of One-Tailed t Test Conducted at .001 Level of Significance.	154
70	Ultimate Robustness of Two-Tailed t Test Conducted at .001 Level of Significance.	155
71	Relationship between Population Moments (especially the Third) and the Robustness of the Sample Mean at N = 2 . . .	282
72	Relationship between Population Moments (especially the Third) and the Robustness of the Sample Mean at N = 16. . .	283
73	Relationship between Population Moments (especially the Third) and the Robustness of the Sample Mean (averaged over ten values of N).	284

STUDIES IN RESEARCH METHODOLOGY

VI. The Central Limit Effect for a Variety of Populations and the Robustness of Z, t, and F

James V. Bradley, PhD

INTRODUCTION

It is often claimed that certain parametric test statistics are "robust", by which is meant that their probabilities of Type I and Type II errors (especially the former) tend to remain about the same when an assumption (the one against which they are allegedly robust) is violated as they are when the assumption is true. Thus if the true probability of false rejection, i.e., a Type I error, is α when all assumptions are met, it will be $\alpha + \epsilon$ when one or more assumptions are violated. The test acquires a reputation of robustness if a sufficiently well-regarded statistical Authority considers ϵ to be negligible. However, ϵ is not a constant for a given violation but varies with sample size and significance level (among other things) neither of which is mentioned in the statement of the assumption (and neither of which is generally mentioned in the allegation of robustness!). These points were brought out in an earlier report of this series (ref. 4), which investigated the robustness of certain one-sample parametric test statistics against violation of the normality assumption alone (and in which the robustness of Z reflected the Central Limit Effect as it will be shown to do in the present study also). The results of this earlier study were such as to prove extremely discomforting to research workers relying upon a reputed robustness to insure the approximate validity of their tests. The question, however, arises as to whether or not these results are peculiar to one-sample tests. Had the two-sample analogues of these same one-sample tests been investigated under precisely the same violation of assumptions, i.e., under precisely the same "degree" of nonnormality, perhaps they would have proven far more robust and entirely acceptable. Or, perhaps the drastic nonrobustness found is peculiar to violation of the assumption of normality and would not be found in the case of, say, heterogeneity of variance. Or, finally, perhaps the previous results are peculiar to the particular type and "degree" of nonnormality investigated, i.e., are attributable to the unique contour of the population distribution. The present study was undertaken to provide an answer to these questions. Indeed, the present study may be regarded as simply a "continuation" of the earlier study: the populations sampled are the same populations sampled earlier, or are functions or combinations of them; the sampling procedure is the same; the statistical controls are essentially the same; the number of samples drawn is the same and the sample sizes are analogous; the test statistics investigated are the multi-sample analogues of the one-sample statistics investigated earlier; and, finally, the tabular presentation of data is the same. One more thing is the same - the findings - extremely damaging evidence against unqualified allegations of robustness.

POPULATIONS SAMPLED

Four populations were investigated. They are designated X, Y, A, and B. The X and Y populations are identical to those investigated in the earlier study (ref. 4). The A and B populations are identical respectively to the X and Y populations, except for a change of the scale parameter, i.e., variance.

The X population is asymmetrical, discrete and finite. It consists of 100,000 units all of which are integers in the range from 89 to 218. Its mean is $\mu = 100$ and its variance is $\sigma^2 = 125.70684$. It is a composite population formed by combining three components: (1) 90,000 integers with mean 96.5 and standard deviation 1.770, (2) 9,500 integers with mean 130 and standard deviation 8.340 and (3) 500 integers with mean 160 and standard deviation 18.730. The three component populations are all exactly symmetrical and all have a shape which is as nearly normal as it was possible to construct with the number of units available and with units having only integral values. The composite population is a very slightly distorted representation of the shape of an actual population of time scores for the operation of a push button by a single subject. The first component corresponds to errorless trials, the second to trials in which the button was missed on the first thrust of the forefinger, but operated on the second thrust, and the third component to trials in which there were exactly two misses before a successful hit.

The Y population is symmetrical, discrete and finite. It consists of 100,000 units with integral values ranging from 50 to 150. Its mean is $\mu = 100$ and its variance is $\sigma^2 = 125.70684$, i.e., its mean and variance are exactly the same as those of the X population. The Y population is as nearly normal in shape as it was possible to construct a population of 100,000 integers with variance as given above.

The A population is asymmetrical, discrete and finite. It consists of 100,000 units having only integral values or values halfway between adjacent integers and ranging in value from 94.5 to 159. Its mean is $\mu = 100$ and its variance is $\sigma^2/4 = 31.42671$. The A population was obtained from the X population by simply moving each X value halfway to the mean, and calling the "relocated" X value an A value. Thus $A_i = 100 + (1/2)(X_i - 100)$, and A_i has exactly the same ordinate, i.e., frequency, as X_i. The A population, therefore, is "identical" to the X population except that it has 1/4 the variance.

The B population was obtained from the Y population in the same way that the A population was derived from the X population. It consists of 100,000 units assuming integral values or values halfway between adjacent integers in the range from 75 to 125, has mean $\mu = 100$ and variance $\sigma^2/4 = 31.42671$. It was obtained by the transformation $B_i = 100 + (1/2)(Y_i - 100)$. It is quasi-normal to exactly the same degree as the Y population.

Figure 1 shows the X and Y populations. The A and B populations, respectively, are "identical" to them except that they have variances only one-fourth as large. By merely performing a linear transformation upon the abscissa <u>scale</u>, therefore, this same figure 1 can be made to depict the A and B populations.

The Y population is "normal" except for its discreteness and finite range. The previous study showed that its discreteness caused appreciable distortions from normal-theory in the sampling distributions of the statistics investigated

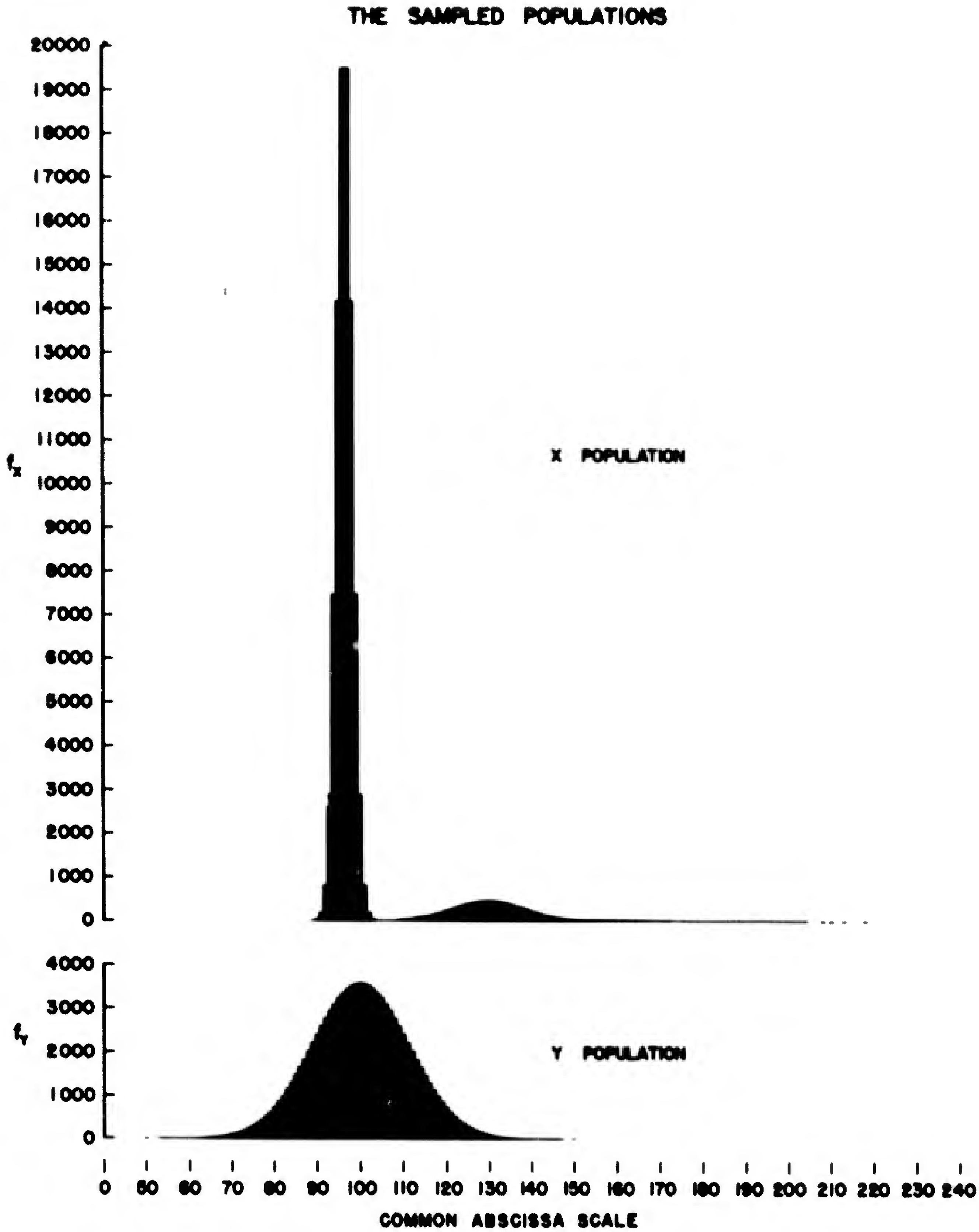

Figure 1. The X and Y Populations

when sample size, N, was 2. At values of N above 4, no such distortions were detected. Results of the present study suggest that discreteness may be responsible for detectable departures from normal theory when N is as large as 8. However, the distortions at N = 4 and N = 8 are quite mild relative to those appearing at N = 2, and neither study detected any distortions attributable to discreteness above an N of 8. Therefore, when N exceeds 8 and all samples are drawn from the Y population, the statistics based upon them should have essentially their normal-theory distributions. Consequently these empirical sampling distributions will serve as controls, i.e., will serve as checks upon the randomness of sampling, the programming of the computer, and the general accuracy of the data and formulae used (i.e., the absence of human error unrelated to the programming of the computer).

The effect of violating the normality assumption alone can be studied by drawing all samples from the X population (homogeneous shapes) or by drawing some samples from the X population and all others from the Y population (heterogeneous shapes).

The effect of violating only the assumption of homogeneity of variance can be studied by drawing some samples from the Y population and all others from the B population.

The combined effect of violating both the assumption of normality and that of homogeneity of variance (if it <u>is</u> an assumption) can be studied by drawing samples from both the X and A populations (homogeneous nonnormal shapes, heterogeneous variances), or from both the X and B or from both the Y and A populations (heterogeneous shapes, heterogeneous variances).

STATISTICS INVESTIGATED

When the tested hypothesis and all assumptions are true, all of the test statistics investigated in this study have distributions belonging to one of three general types: Z, t, or F. However, since seven different formulae are required to identify, and calculate, the statistics, it is better to think of them as belonging to seven general categories rather than three. The seven tests are: (1) the one-sample Z test for an hypothesized population mean, (2) the one-sample t test for an hypothesized population mean, (3) the two-sample t test, using matched pairs of observations, for an hypothesized difference between population means, (4) the two-sample Z test for an hypothesized difference between population means, (5) the two-sample t test, (with unmatched observations) for an hypothesized difference between population means, (6) the two-sample F test for an hypothesized difference between population variances, and (7) the (multi-sample) analysis-of-variance F test for an hypothesized difference among the means of several populations. The formulas by which these test statistics are calculated are given below. The formulas are for the null case where the hypothesis tested is that there is "no difference", since this was always the case treated in this study. The symbols used in these formulas do not always coincide with those used in this study. They are the standard ones found in statistical texts and are therefore not defined. (In case of ambiguity, the reader may refer to appendix I where the formulas actually used are given with explicit definition of terms and are expressed in the symbols peculiar to this study.) The formulas, in the same order in which the tests were identified above, are:

(1) $$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$$

(2) $$t = \frac{\bar{X} - \mu}{\sqrt{\frac{\sum (X - \bar{X})^2}{N(N - 1)}}} \qquad df = N - 1$$

(3) $$t = \frac{\bar{D}}{\sqrt{\frac{\sum (D - \bar{D})^2}{N(N - 1)}}} \qquad df = N - 1$$

(4) $$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

(5) $$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sum (X_1 - \bar{X}_1)^2 + \sum (X_2 - \bar{X}_2)^2}{N_1 + N_2 - 2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}} \qquad df = N_1 + N_2 - 2$$

$$(6) \qquad F = \frac{\dfrac{\sum (X_1 - \bar{X}_1)^2}{N_1 - 1}}{\dfrac{\sum (X_2 - \bar{X}_2)^2}{N_2 - 1}} \qquad df = N_1 - 1,\ N_2 - 1$$

$$(7) \qquad F = \frac{\dfrac{\sum_{i=1}^{k} N_i \, (\bar{X}_i - \bar{\bar{X}})^2}{k - 1}}{\dfrac{\sum_{i=1}^{k} \sum_{j=1}^{N_i} (X_{ij} - \bar{X}_i)^2}{\sum_{i=1}^{k} (N_i - 1)}} \qquad df = k - 1,\ \sum (N_i - 1)$$

In this study, the Z tests will be represented by the conventional Z, the t tests by a T (since the computer only prints capital letters), the F test for variances (6) by an R (standing for Ratio) and the analysis-of-variance F test for means (7) by the usual F.

All of the above tests assume that all the sampled populations are normally distributed. Tests (5) and (7), the two-sample t test using unmatched observations and the analysis-of-variance F test, also assume that all sampled populations have the same variance. Homogeneity of variance is not assumed by the one-sample tests (1) and (2) where the fact that only a single population is involved obviates the question of heterogeneity. Nor is it assumed by (3), the two-sample t test using matched observations. This test uses only a single variance, that of the difference-scores. Furthermore, it can be regarded as a one-sample test which draws observations from a "population" of difference-scores and tests the hypothesis that the mean of the difference-score population has a specified value (zero in the null case). Therefore, as a one-sample test, only one variance is involved - the variance of the "population" of difference-scores. The two-sample Z test, (4), does not assume homogeneity of variance, but, instead, takes explicit account of any heterogeneity by substituting the exact values of the population variances into its formula. Finally, (6) does not "assume" homogeneity of variance because homogeneity of variance is what it is testing.

The above formulae merely identify the seven tests which were investigated. They do not specify the particular conditions (i.e., populations sampled and relative sample sizes), under which the investigation occurred. Henceforth, the terms "test", "statistic" or "test statistic" will imply the entire set of sampling conditions under which a test statistic was investigated, except for the value of a parameter, N, related to absolute sample size, and appearing as a variable in the definition of a test statistic. In all, 122 such sets or combinations of conditions were investigated at each of ten value of N and for each such combination of conditions an empirical sampling distribution of 10,000 values of the test statistic was obtained. Therefore, there are 122 "test statistics", as redefined, rather than seven. These are the basic test statistics. Data relevant to the empirical sampling distributions of all of them are given in the appendix. Some of these statistics are equivalent in the sense that their respective empirical sampling distributions are simply independent replications of the sampling distribution for a single statistic or in the sense that their true distributions are mirror-images of each other. In such cases, the equivalent statistics and their individual sampling distributions are represented in the body of the report by a single test statistic whose empirical sampling distribution was obtained by combination from the distributions of the individual, equivalent, statistics. After all such combinations, and after omitting other distributions, such as those obtained solely for the purpose of detecting nonrandomness of sampling or errors of programming, 81 statistics and their distributions remain which are dealt with, graphically, in the body of the report. Because of the large number of test statistics involved, they will be identified by means of a simple code which indicates the type of statistic involved, the population from which each sample was drawn, and the respective sizes of the samples. This code will be explained in a later section.

SAMPLING SCHEME

Only the broad outline of sampling will be given here. The more technical details of sampling and a description of the elaborate checks upon the accuracy of the machine program and the randomness of sampling (all of which were very satisfactorily passed by the final, reported, product) have been relegated to the Appendix.

Sampling was performed by an IBM 7090 electronic digital computer. For each of the following values of N, 2, 4, 8, 16, 32, 64, 128, 256, 512, and 1024, the computer generated a sequence of 10,000(4N) random numbers (actually "pseudo-random" numbers). Thus, for each N value, the computer generated a sequence of 10,000 blocks of random numbers, each block containing a sequence of 4N random numbers. These random numbers were used to identify and draw from their respective populations the observations required by the various samples contributing to a test statistic. No statistic investigated in the study proper is ever based upon more than 4N observations in all, i.e., counting all samples. Therefore, a single block, or sequence, of 4N random numbers contained enough random numbers to identify all of the observations in all of the samples required for the calculation of a single value of any one of the test statistics investigated. Actually, each such block of 4N random numbers was used by all 122 of the test statistics, i.e., every one of the test statistics investigated used the same pool of 4N random numbers exactly once to obtain the observations required for the calculation of a single value of the test statistic. Surplus numbers in the pool, i.e., those not required by a given statistic, were simply discarded so far as that statistic was concerned, and the "next" value of that statistic was obtained using the next block of 4N entirely different and nonoverlapping random numbers. Since sample sizes were always integral multiples of N, it was logical to divide the pool of 4N random numbers into 4 sets of N numbers each and use either all or none of the numbers comprising a set to identify the observations to be selected for a sample. Finally, it was convenient to let these four sets be determined by the sequence in which the numbers were generated, so that the four sets are simply the first N, second N, third N, and fourth N numbers in the sequence of 4N numbers as generated.

To the extent that they use the same subsets of random numbers, the various test statistics will be equally affected by that portion of "chance" which can be ascribed to randomness. (This is discussed in detail in the earlier report (ref. 4, pp 48-52, 59-67).) Thus maximum comparability between sampling distributions of the various statistics would seem to require maximum overlap in the random numbers used. For this reason, in the case of two-sample tests, one sample was built up by using random numbers in sequence, starting from the beginning of a sequence of 4N numbers, and the other sample was built up by starting at the end of the block of 4N random numbers and working backward in reverse sequence. An analogous procedure was followed for three- and four-sample tests (with a single exception which appears only in the appendix).

Ten thousand blocks of 4N random numbers were generated at each value of N, and for each basic test statistic, a value of the statistic was obtained for each block. Therefore at each value of N an empirical distribution of 10,000 values of each basic test statistic was obtained. A different 10,000 blocks of 4N random numbers was generated for each of the ten different values of N investigated. Therefore, for a given test statistic the ten different sampling distributions corresponding to the ten values of $N \leq 1024$ are statistically independent. This contrasts with the fact that the sampling distributions of different test statistics at the same value of N are usually nonindependent.

IDENTIFYING CODES FOR TEST STATISTICS

The basic one-sample statistics dealt with in this study are identified by a code of the form

SP K

The first letter, S, stands for the general type of statistic involved (of which there were two, Z and T). The second letter, P, stands for the population (either X or Y) from which observations were drawn. And K is a number indicating which of the four subsets of N random numbers, within a block of 4N, were used to draw the required observations. See table I, which lists the codes and sampling scheme for all one-sample statistics dealt with in this study. For given values of S and P, SP 1, SP 2, SP 3, and SP 4 are simply four independent replications of the same one-sample statistic, each having an empirical sampling distribution of 10,000 values. The earlier study (ref. 4) yields a fifth such independent replication. Its statistic, which is relabeled SP 0, is also based upon N observations and has an empirical sampling distribution of 10,000 values. By simply combining these five independent empirical sampling distributions of 10,000 values each, one obtains an empirical sampling distribution of 50,000 values for the same test statistic, which is now designated simply SP. While SP 1, SP 2, SP 3, SP 4, SP 0, and SP are each based upon samples of size N, i.e., containing N observations each, the statistic SP 1234, which was also investigated, is based upon samples of size 4N. Therefore, when N = 512 and N = 1024, SP 1234 is equivalent to, and provides a "continuation" of, SP at respective sample sizes of 2048 and 4096, although its empirical sampling distributions in these cases contain only 10,000, rather than 50,000, values, and are not independent of the distributions of SP at N = 512 and N = 1024, respectively.

TABLE I

SAMPLING SCHEME AND IDENTIFYING CODES FOR ONE-SAMPLE STATISTICS

Random Numbers Used to Draw N Observations from Population Indicated by Cell Entry: 1st N	2nd N	3rd N	4th N	Previous Study	Z Statistics	T Statistics	Size of Each Statistic's Sampling Distribution	Z Statistics	T Statistics	Size of Each Statistic's Sampling Distribution
Y					ZY 1	TY 1	10,000			
	Y				ZY 2	TY 2	10,000			
		Y			ZY 3	TY 3	10,000	ZY	TY	50,000
			Y		ZY 4	TY 4	10,000			
				Y	ZY 0	TY 0	10,000			
Y	Y	Y	Y		ZY 1234	TY 1234	10,000	ZY (When "N" = 4N = 2048 or 4096)	TY	10,000
X					ZX 1	TX 1	10,000			
	X				ZX 2	TX 2	10,000			
		X			ZX 3	TX 3	10,000	ZX	TX	50,000
			X		ZX 4	TX 4	10,000			
				X	ZX 0	TX 0	10,000			
X	X	X	X		ZX 1234	TX 1234	10,000	ZX (When "N" = 4N = 2048 or 4096)	TX	10,000

Note: Statistics based entirely upon samples from the Y population were obtained only to provide "control" distributions. The ZY 1, ZY 2, ZY 3, and ZY 4 statistics provide the primary control distributions in that they reflect the randomness or nonrandomness of the basic building blocks considered individually (i.e., their distributions reflect the randomness or nonrandomness of the 10,000 sets of "1st N", "2nd N", "3rd N", and "4th N" random numbers, respectively.)

If we let P stand for an observation drawn from the P population as well as for the population itself, the generic formulas for the one-sample statistics of interest are:

$$ZP = \frac{\bar{P} - \mu}{\sqrt{\frac{\sigma_P^2}{N}}} \quad \text{and} \quad TP = \frac{\bar{P} - \mu}{\sqrt{\frac{\sum (P-\bar{P})^2}{N\ (N-1)}}} .$$

When P = Y, the statistics of major interest are ZY 1, ZY 2, ZY 3, and ZY 4. These statistics provide the primary control distributions used to determine whether or not the computer output at a given N value was acceptable, in that they reflect the randomness or nonrandomness of the individual "building blocks," the 1st N, 2nd N, 3rd N, and 4th N random numbers, respectively, within the fundamental sequence of 4N random numbers. Since the N random numbers in each of these building blocks were always used in their entirety, if at all, being applied exclusively to a single sample for a given statistic, it would be particularly useful to base tests of randomness upon the individual building blocks. Instead of merely indicating that the entire set of observations upon which a statistic is based are sufficiently random <u>in toto</u>, such tests would provide information as to the randomness or nonrandomness of each of the component samples used by the test statistic. The Y population is exactly symmetric and as nearly normal as its discreteness will permit, and the ZY statistics are simply linear transformations upon the means of samples of observations drawn from it. Therefore, if sampling is truly random the true sampling distributions of the ZY's will be symmetric for all sample sizes and, as the earlier study showed, except at very small sample sizes their empirical sampling distributions will be indistinguishable from the normal by a specified chi-square test of fit. Therefore the randomness of each of the four individual sets of 10,000 building blocks can be tested indirectly by means of chi-square tests of goodness of fit to normality applied to the four corresponding ZY distributions and supplemented by tests of symmetry at small sample sizes when the tests of normality fail.

Actually any statistic based exclusively upon samples from the Y population provides a control distribution, and a variety of such statistics were obtained for purposes of control. However, their distributions are generally not mutually independent unless nonoverlapping, i.e., independent, sets of random numbers were used to draw the observations upon which their test statistic is based. Thus while a variety of nonindependent "control" distributions is desirable in that they indicate the degree of control for the maximally analogous "experimental" distributions, so far as acceptance or rejection of an entire batch of computer output is concerned the decision should be determined by independent tests. The four ZY statistics provide the distributions for four such independent, and therefore "basic," tests of randomness.

When P = X, the statistics of primary interest are the "overall" ZX and TX statistics, having distributions of 50,000 values at each $N \leq 1024$ and distributions containing 10,000 values each at "N" = 2048 and 4096. The ZX 1, ZX 2, ZX 3, and ZX 4 distributions are simp four independent replications of the Z_x distribution obtained in the earlier study and relabled ZX 0 here. By combining all five distributions of 10,000 values each into a single distribution of 50,000 values, one simply obtains greater reliability, i.e., the distribution of 50,000 values would be expected to approximate the true distribution of $Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{N}}}$ more closely than would any of the distributions of 10,000 values. (A completely analogous statement can be made for the one-sample T statistics.) The gain, therefore, is in precision of estimate, and, indeed, the empirical distributions of ZX and TX, containing 50,000 values each, should differ only very slightly from the true distributions, except at the extreme tails. Before being combined with data from the present study, the data for both Z and T from the previous study were reanalyzed in terms of the more accurate interval boundaries used in the present study (see appendix II), i.e., empirical cumulative probabilities were obtained by cumulating empirical values up to more accurately determined boundaries, thus obtaining completely homogeneous and consistent treatment of data from the two studies. The boundary values used in the previous study were highly accurate for Z, and the reanalysis resulted in no changes; however, some of the previous t boundaries had been obtained by interpolation and a few small changes resulted from the reanalysis in some of these cases.

All but one of the two-sample statistics investigated in this study are identified by a code of the form

$$SPQ\ C_1NC_2N$$

The first letter, S, stands for the symbol identifying the type of two-sample statistic involved. In specific cases it will be replaced by one of the letters Z, T, or R. The second and third letters, P and Q, stand for the populations from which the first and second samples, respectively, were drawn. Thus in specific cases X, Y, A, or B (or D in a single exceptional case which is explained later) will occupy the positions of the letters P and Q in the generic formula above. When all samples are drawn from the same population, P = Q and the Q is sometimes omitted from the code. C_1N is the size of the first sample, i.e., the sample drawn from the first-listed, P, population, and C_2N is the size of the second sample, the sample drawn from the second-listed, Q, population. C_1 and C_2 are integers having one of the values 1, 2, or 3. When C_1 or C_2 equals 1, it is omitted leaving only the N.

The single exception to the above code occurs with the two-sample t test for matched pairs. This test is investigated under only one set of sampling conditions (other than absolute sample size): 2N observations are drawn from the X population and another 2N are drawn from the Y population. In order to produce correlated observations, the same 2N random numbers are used to draw both samples. The resulting statistic would be identified as TXY 2N2N except that this designation is required for the t test using unmatched observations. However, one can regard the two-sample t test for matched pairs as a one-sample test in which 2N observations have been drawn from a population of X-Y difference-scores, where D = X-Y. Therefore, the matched pair t test was coded, more or less, as a one-sample test and identified as TD 2N.

If we let P, Q, and D stand for observations drawn from the P, Q, and D populations, as well as for the populations themselves, then the formulas for the four general types of two-sample test statistics investigated in this study can be related to their identifying codes as follows.

$$ZPQ\ C_1NC_2N = \frac{\bar{P} - \bar{Q}}{\sqrt{\frac{\sigma_P^2}{C_1N} + \frac{\sigma_Q^2}{C_2N}}}$$

$$TPQ\ C_1NC_2N = \frac{\bar{P} - \bar{Q}}{\sqrt{\dfrac{\sum_1^{C_1N}(P-\bar{P})^2 + \sum_1^{C_2N}(Q-\bar{Q})^2}{C_1N + C_2N - 2}\left(\dfrac{1}{C_1N} + \dfrac{1}{C_2N}\right)}}$$

$$TD\ 2N = \frac{\bar{D}}{\sqrt{\dfrac{\sum_1^{2N}(D-\bar{D})^2}{2N\ (2N-1)}}}$$

$$RPQ\ C_1NC_2N = \frac{\dfrac{\sum_1^{C_1N}(P-\bar{P})^2}{C_1N - 1}}{\dfrac{\sum_1^{C_2N}(Q-\bar{Q})^2}{C_2N - 1}}$$

Table II lists all of the two-sample statistics investigated in this study and shows what subsets of random numbers (within each basic block of 4N random numbers) were used to identify the observations to be drawn for each sample. For each test statistic two codes are listed, the one used in the appendix and the one used in the body of this report. Both codes are consistent with the definitions given above. When the two codes differ it is only in regard to which of the two samples is to be designated as the "first" sample, or, equivalently, it is only with regard to which of the two sampled populations is to be listed first. The code used in the appendix follows the convention that the first sample is always the larger sample, when sample sizes differ, and is drawn by using the first C_1N random numbers within a block of 4N to identify the required C_1N observations from the P population. (The second sample is obtained by using the last C_2N random numbers to draw the necessary C_2N observations from the Q population.) As a result, the first-listed population is always the one contributing the larger sample and P and Q always correspond "spatially" with the subsets of random numbers used to sample from them. Statistics whose analogous

TABLE II

Sampling Scheme and Identifying Codes for Two-Sample Statistics

Code Used in Body of Report			Random Numbers (Column Headings) used by Statistic to Draw Observations from the Population Indicated by Cell Entry				Code Used in Appendix		
Z Statistics	T Statistics	(Footnotes)	1st N	2nd N	3rd N	4th N	Z Statistics	T Statistics	(Footnotes)
			Y			Y	ZY NN	TY NN	S
			Y	Y		Y	ZYY 2NN	TYY 2NN	S
			Y	Y	Y	Y	ZYY 3NN	TYY 3NN	S
			Y	Y	Y	Y	ZYY 2N2N	TYY 2N2N	S
ZX NN	TX NN	f	X			X	ZX NN	TX NN	S
ZXX 2NN	TXX 2NN		X	X		X	ZXX 2NN	TXX 2NN	
ZXX 3NN	TXX 3NN		X	X	X	X	ZXX 3NN	TXX 3NN	
ZXX 2N2N	TXX 2N2N	f	X	X	X	X	ZXX 2N2N	TXX 2N2N	S
ZXY 2NN	TXY 2NN		X	X		Y	ZXY 2NN	TXY 2NN	
ZXY N2N	TXY N2N		Y	Y		X	ZYX 2NN	TYX 2NN	
ZXY 3NN	TXY 3NN		X	X	X	Y	ZXY 3NN	TXY 3NN	
ZXY N3N	TXY N3N		Y	Y	Y	X	ZYX 3NN	TYX 3NN	
ZXY 2N2N	TXY 2N2N	c [	X	X	Y	Y	ZXY 2N2N	TXY 2N2N	]
			Y	Y	X	X	ZYX 2N2N	TYX 2N2N	
	TYB 2NN	f	Y	Y		B		TYB 2NN	S
	TYB N2N	f	B	B		Y		TBY 2NN	S
	TYB 3NN	f	Y	Y	Y	B		TYB 3NN	S
	TYB N3N	f	B	B	B	Y		TBY 3NN	S
	TYB 2N2N	fc [	Y	Y	B	B		TYB 2N2N	S]
			B	B	Y	Y		TBY 2N2N	S
ZXA 2NN	TXA 2NN		X	X		A	ZXA 2NN	TXA 2NN	
ZXA N2N	TXA N2N		A	A		X	ZAX 2NN	TAX 2NN	
ZXA 3NN	TXA 3NN		X	X	X	A	ZXA 3NN	TXA 3NN	
ZXA N3N	TXA N3N		A	A	A	X	ZAX 3NN	TAX 3NN	
ZXA 2N2N	TXA 2N2N	c [	X	X	A	A	ZXA 2N2N	TXA 2N2N	]
			A	A	X	X	ZAX 2N2N	TAX 2N2N	
ZXB 2NN	TXB 2NN		X	X		B	ZXB 2NN	TXB 2NN	
ZXB N2N	TXB N2N		B	B		X	ZBX 2NN	TBX 2NN	
ZXB 3NN	TXB 3NN		X	X	X	B	ZXB 3NN	TXB 3NN	
ZXB N3N	TXB N3N		B	B	B	X	ZBX 3NN	TBX 3NN	
ZXB 2N2N	TXB 2N2N	c [	X	X	B	B	ZXB 2N2N	TXB 2N2N	]
			B	B	X	X	ZBX 2N2N	TBX 2N2N	
ZYA 2NN	TYA 2NN		Y	Y		A	ZYA 2NN	TYA 2NN	
ZYA N2N	TYA N2N		A	A		Y	ZAY 2NN	TAY 2NN	
ZYA 3NN	TYA 3NN		Y	Y	Y	A	ZYA 3NN	TYA 3NN	
ZYA N3N	TYA N3N		A	A	A	Y	ZAY 3NN	TAY 3NN	
ZYA 2N2N	TYA 2N2N	c [	Y	Y	A	A	ZYA 2N2N	TYA 2N2N	]
			A	A	Y	Y	ZAY 2N2N	TAY 2N2N	
	TD 2N		X & Y			X & Y		TD 2N	
R Statistics							**R Statistics**		
			Y	Y	Y	Y	RY 2N2N		
RXX 2N2N			X	X	X	X	RX 2N2N		
RXY 2N2N			X	X	Y	Y	RXY 2N2N		

S True distribution of the test statistic (for both Z and T) is symmetric about zero.

] Enclosed test statistics of the same type (for both Z and T) have true distributions which are mirror-images of each other (about an axis of zero).

c [Data from the two mirror-image distributions were combined into a single empirical distribution of 20,000 values.

f Data from the left and right tails of the symmetric distribution were combined in graphing results (i.e., empirical distribution of k values was "folded over" about an axis through zero to yield k values for a single tail).

samples were obtained using the same random numbers tend to be maximally comparable in that their empirical distributions are distorted by maximally correlated chance effects. The code convention followed in the appendix facilitates such comparisons. The code used in the body of the report follows the convention that for a given pair of populations one member is always the first-listed population. Thus, while in the appendix code, for example, some statistics are listed as TXA and have the numerator $\bar{X}-\bar{A}$ and others are identified as TAX and have the numerator $\bar{A}-\bar{X}$, in the code used in the body of the report all of the corresponding statistics are listed as TXA and have the numerator $\bar{X}-\bar{A}$. If one were interested in the effect of relative sample size, in the former case one would have to compare the <u>left</u> tails of TXA 3NN, TXA 2NN, and TXA 2N2N with the <u>right</u> tails of TAX 2NN and TAX 3NN; in the latter case one could simply compare the left tails of TXA 3NN, TXA 2NN, TXA 2N2N, TXA N2N, and TXA N3N. The latter approach is particularly desirable when results are graphed as is the case in the body of this report.

To each of the codes listed on the right of table II, i.e., to each of the "appendix" codes, there corresponds an empirical distribution of 10,000 values of the test statistic which that code identifies. (This is also true of table III.) These are the basic data. The sampling distributions treated in the body of the report are derived from them in such a way as to (a) give all Z or T statistics based on the same pair of populations comparable numerators, i.e., numerators in which the subtracted mean always represents the same population, (b) take advantage of any increment of "information" by which precision of estimate might be increased, irrespective of whether or not the various sources of information are entirely independent. The following four paragraphs explain in detail both the rationale and the method by which this was accomplished.

From the formulae and definitions already given it follows that

ZLM JNKN = -ZML KNJN and TLM JNKN = -TML KNJN provided that the statistics on both sides of the equality sign are based upon exactly the same set of JN observations from the L population and exactly the same set of KN observations from the M population (in which case $\bar{L}-\bar{M} = -(\bar{M}-\bar{L})$, i.e., the numerator of ZLM JNKN is simply the negative of the numerator of ZML KNJN, and likewise for the corresponding T statistics). It is clear, therefore, that the <u>true</u> distribution of ZLM JNKN must be the mirror image of the true distribution of ZML KNJN, (and likewise for the corresponding T's). Consequently, for example, TXA N3N = - TAX 3NN and if one wishes to convert the TAX 3NN distribution into a TXA N3N distribution one need only rotate the former through 180 degrees about an axis through zero, i.e., one need only regard the left tail of the TAX 3NN distribution as the right tail of the TXA N3N distribution (and vice versa) and obtain cumulative probabilities accordingly. By following this procedure statistics whose code lists populations in the reverse of the desired order, and their distributions, were converted into the statistics and distributions required in the body of the report.

When sample sizes are unequal, such as is the case for TAX 3NN, there is a single empirical distribution of 10,000 values, which can be converted into its mirror image, e.g., the distribution appropriate to TXA N3N, if desired, and the latter corresponds to no other distribution. However, when sample sizes are equal, two empirical Z or T distributions were obtained (for cases in which samples were drawn from two different populations). For example, empirical distributions of 10,000 values each were obtained for both TXA 2N2N and TAX 2N2N. The <u>true</u> sampling distributions of these two statistics are exact mirror images of each other, about an axis of zero, because the formula for one statistic is simply the negative of the formula for the other. But their <u>empirical</u> distributions are <u>not exact</u> mirror images. Although a single value of each statistic is obtained using the same pool of 4N random numbers, the X observations are identified by the first 2N random numbers in the former case but by the second 2N random numbers in the latter, so the corresponding single empirical value of TXA 2N2N does not have to be simply the negative of the single empirical value of TAX 2N2N. Therefore the empirical distribution of 10,000 independent values of TAX 2N2N can be rotated 180 degrees about an axis through zero and combined with the empirical distribution of 10,000 independent values of TXA 2N2N to yield an empirical distribution of 20,000 partially "dependent" values of TXA 2N2N. The effect of this "criss-cross" combining is, in essence, to calculate TXA 2N2N twice for each pool of 4N random numbers used, once using the first 2N numbers to identify the X's and the second 2N to identify the A's, and once using the second 2N numbers to draw the X's and the first 2N to obtain the A's. The resulting distribution is not the same as would be obtained using 20,000 different sets of 4N random numbers and using each set only once; however, because of the counterbalancing of chance effects, the net result of the "criss-cross" procedure upon the distribution of 20,000 partially "dependent" values of TXA 2N2N may be to produce greater, rather than less, precision of estimate of the true distribution as contrasted with the precision of estimate obtained from a distribution of 20,000 completely independent values. (Because of the "dependence" of chance effects, however, one should not attempt to apply conventional statistical analysis, e.g., confidence or tolerance limits, to such distributions, i.e., those obtained by criss-cross combination, since such analysis presupposes completely independent sampling.)

The true distribution of a two-sample Z or T statistic must be symmetric about zero if either (a) both sampled populations are symmetric about the same point, or (b) the two sampled populations are identical and the two samples are of equal size. This can be seen as follows. If a population, say P, is symmetric about μ, then for every deviation $d = (P-\mu)$ the opposite deviation $-d = -(P-\mu)$ is equally probable. For every sample of deviations, there will be an "opposite" sample of deviations having the same absolute magnitudes but opposite algebraic signs, and the two samples will be equally probable. The two samples will have means of the same absolute magnitude but opposite algebraic sign, i.e., $\bar{d} = (\bar{P}-\mu)$ and $-\bar{d} = -(\bar{P}-\mu)$ and since $\{d-\bar{d}\}^2 = \{-d-(-\bar{d})\}^2$, will have identical sums of squares (furthermore since $\{d-\bar{d}\}^2 = \{(P-\mu)-(\bar{P}-\mu)\}^2 = \{P-\bar{P}\}^2$ the sums of squares is the same whether original observations or their deviations are used). Likewise, for sampling from the Q population mutually "opposite" samples of deviations are equally likely and have identical sums of squares and means of $(\bar{Q}-\mu)$ and $-(\bar{Q}-\mu)$ respectively. Since the two P samples are equally probable and the two Q samples are equally probable, any pairing of one of the two P samples with one of the two Q samples is equally probable. Therefore the pair of samples yielding the difference $(\bar{P}-\mu)-(\bar{Q}-\mu) = \bar{P}-\bar{Q}$ has the same probability as the pair yielding the difference $-(\bar{P}-\mu) + (\bar{Q}-\mu) = -(\bar{P}-\bar{Q})$, and the total sum of squares for both pairs of samples is the same. Therefore for every pair of samples yielding a given numerator for Z or t and a given denominator for t, there is an equally probable pair of samples yielding a numerator of the same absolute magnitude but opposite algebraic sign and yielding the same denominator. It follows that the true distributions of Z and t must be symmetric about zero. Now turning to the second case, if the two sampled populations are identical (irrespective of whether they are symmetric or not) and samples are of equal size, then the "first" and "second" samples differ (on an *a priori* basis) only as to sequence or some such arbitrary label. Therefore the observations drawn for one sample were just as likely to have been drawn for the other. This is equivalent to saying that the numerator $\overline{P}-\overline{Q}$ is exactly as likely as the numerator $\overline{Q}-\overline{P}$. But the latter can be written $-(\overline{P}-\overline{Q})$, i.e., as the negative of the former. Since the two sums of squares are added in the denominator, irrespective of "sequence", the denominators for the two cases are the same. Again it follows that the true distributions of Z and t must be symmetric about zero.

The true sampling distributions of Z and T are therefore symmetric about an axis through zero when the two sampled populations are Y and Y or Y and B or when samples are of equal size and the "two" (identical) sampled populations are X and X. In these cases, the negative portion of the empirical sampling distribution may be regarded (by ignoring or reversing algebraic sign) as a virtually independent replication of the positive portion. Better still, both portions may be regarded as independent, and therefore summable, empirical frequency distributions of $|Z|$ or $|T|$ as the case may be. In any case, precision of estimate can be greatly increased by rotating the negative portion of the empirical distribution 180 degrees about an axis through zero and combining its frequencies with those of the positive portion. This "folding over" produces a "half distribution" containing the same number of values as previously spread out over an entire distribution. The effect is to double the number of values used by the empirical distribution to estimate a symmetric half of the true distribution. The data may then be presented as applying to a single tail of the distribution, or, the mirror image of the empirical "half distribution" may be used to represent the opposite tail. Both procedures were followed in this study depending upon what was necessary to preserve local consistency in presentation of results. In all but one case folding over produced a "half distribution" of 10,000 values. The one exception was TYB 2N2N which, before folding, already consisted of 20,000 values obtained by the "criss-cross" combination of TYB 2N2N with TBY 2N2N. After folding, therefore, TYB 2N2N had an empirical "half distribution" consisting of 20,000 partially dependent values.

The three- and four-sample tests investigated in this study are identified, in the body of the report, by a code of the form

$$SPQ\ C_1N, C_2NC_3NC_4N.$$

As before, S stands for the symbol identifying the type of statistic, which was always the analysis of variance F test. P and Q stand for the populations sampled (there were never more than two). C_1N, C_2N, C_3N, and C_4N are the sizes of the samples drawn. The C's are integers which can take values of 0, 1, or 2. When the test is a three-sample test, one of the C's will be zero, and the entire 0N term is omitted. When a C = 1, the one is omitted in front of the N. The comma in the above code indicates the populations from which the samples were drawn. Sample sizes to the left of the comma pertain to samples drawn from the P population, i.e., the first-listed population; those to the right pertain to samples drawn from the Q population. When P = Q, i.e., when all samples are drawn from a single population, the comma is always omitted, and sometimes the Q is omitted. The code used in the appendix is slightly different. When a CN = 0, the zero is usually included in the code. The comma is always omitted because all but one sample (those whose sizes are listed first) were always drawn from the first-listed population and a single sample of size N (the one listed last) was always drawn from the second-listed population.

Table III lists all of the three- and four-sample test statistics investigated, giving both codes for each statistic, and showing what subsets of random numbers (within each block of 4N random numbers) were used to identify the observations to be drawn for each sample. In analogy with the procedure for two-sample tests, the samples from one population are (with one exception) built up by starting at the beginning of a block of 4N random numbers and using random numbers consecutively in sequence, and the single sample from the other population is always identified by the last N random numbers within a block of 4N random numbers.

TABLE III

SAMPLING SCHEME AND IDENTIFYING CODES FOR THREE- AND FOUR-SAMPLE STATISTICS

Code Used in Body of Report	Random Numbers (Column Headings) Used by Statistic and Populations (Cell Entries) to Which They Were Applied				Code Used in Appendix*
	1st N	2nd N	3rd N	4th N	
	Y		Y	Y	FY NONN
	Y	Y		Y	FYY NNON
	Y	Y	Y	Y	FY 2NNN
	Y	Y	Y	Y	FYY NNNN
	X		X	X	FX NONN
FXX NNN	X	X		X	FXX NNON
FXX 2NNN	X	X	X	X	FX 2NNN
FXX NNNN	X	X	X	X	FXX NNNN
FXY NN,N	X	X		Y	FXY NNON
FXY N,NN	Y	Y		X	FYX NNON
FXY NNN,N	X	X	X	Y	FXY NNNN
FXY N,NNN	Y	Y	Y	X	FYX NNNN
FYB NN,N	Y	Y		B	FYB NNON
FYB N,NN	B	B		Y	FBY NNON
FYB NNN,N	Y	Y	Y	B	FYB NNNN
FYB N,NNN	B	B	B	Y	FBY NNNN
FXA NN,N	X	X		A	FXA NNON
FXA N,NN	A	A		X	FAX NNON
FXA NNN,N	X	X	X	A	FXA NNNN
FXA N,NNN	A	A	A	X	FAX NNNN
FXB NN,N	X	X		B	FXB NNON
FXB N,NN	B	B		X	FBX NNON
FXB NNN,N	X	X	X	B	FXB NNNN
FXB N,NNN	B	B	B	X	FBX NNNN
FYA NN,N	Y	Y		A	FYA NNON
FYA N,NN	A	A		Y	FAY NNON
FYA NNN,N	Y	Y	Y	A	FYA NNNN
FYA N,NNN	A	A	A	Y	FAY NNNN

*The second-listed population, in the appendix code, was always represented by a single sample of size N, drawn using the 4th set of N random numbers in a block of 4N random numbers.

POPULATIONS AND STATISTICS RELEVANT TO THE CENTRAL LIMIT THEOREM

The Central Limit Theorem states that the mean, $\bar{X}$, of a sample of N observations drawn from a (nonnormal) population with mean μ and (finite) variance σ^2 has a distribution which, as N increases, approaches closer and closer to a normal distribution whose mean is μ and whose variance is σ^2/N. The one-sample Z statistic, $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{N}}$, is simply a linear transformation upon the sample mean, and consequently the Central Limit Theorem implies that, under the conditions stated above, the distribution of Z approaches a normal distribution with mean of zero and variance of one, i.e., the "tabled" distribution of the unit normal deviate. The distribution of the one-sample Z statistic is therefore a direct indicator of the Central Limit Effect and is just as informative in that regard as would be the distribution of the sample mean itself.

When the samples are equal in size, the two-sample Z statistic can be regarded as a one-sample Z statistic based upon observations drawn from a population of difference-scores, and therefore its distribution displays the Central Limit Effect for the condition of sampling from such a population. The equivalence can be shown algebraically. Let the two absolute-score populations be designated P and Q, with means μ_P and μ_Q, variances σ_P^2 and σ_Q^2, and observations P_i and Q_i, and let the size of each of the two samples be 2N (to conform to the already established notation to be used in presenting results). Then the two-sample Z statistic, for testing the true hypothesis that the difference between population means is $\mu_P-\mu_Q$ may be written.

$$\text{ZPQ 2N2N} = \frac{(\bar{P}-\bar{Q}) - (\mu_P-\mu_Q)}{\sqrt{\frac{\sigma_P^2}{2N} + \frac{\sigma_Q^2}{2N}}}$$

$$= \frac{\frac{1}{2N}\sum_1^{2N} P_i - \frac{1}{2N}\sum_1^{2N} Q_i - (\mu_P-\mu_Q)}{\sqrt{\frac{\sigma_P^2 + \sigma_Q^2}{2N}}}$$

$$= \frac{\frac{1}{2N}\sum_{1}^{2N} (P_i - Q_i) - (\mu_P - \mu_Q)}{\sqrt{\frac{\sigma_P^2 + \sigma_Q^2}{2N}}}$$

But $\frac{1}{2N}\sum_{1}^{2N} (P_i-Q_i)$ is the mean of the 2N observations comprising a sample of difference-scores, (P_i-Q_i)'s, and can therefore be written $\overline{P-Q}$. Also, the difference between population means, $\mu_P-\mu_Q$, equals the population mean difference, μ_{P-Q}, i.e., the mean of the difference-score populations, for essentially the same reasons that $\bar{P}-\bar{Q} = \overline{P-Q}$, and, finally, since the variance of the difference between two variables equals the sum of their individual variances, $\sigma_P^2 + \sigma_Q^2$ is simply the true variance of the difference-scores, (P_i-Q_i)'s. Upon substituting, therefore,

$$ZPQ\ 2N2N = \frac{\overline{P-Q} - \mu_{P-Q}}{\sqrt{\frac{\sigma^2_{P-Q}}{2N}}}$$

and it is seen that the two-sample Z statistic, ZPQ 2N2N, is equivalently a one-sample Z statistic based on 2N observations drawn from a population of difference-scores, (P_i-Q_i)'s. The distribution of ZPQ 2N2N is therefore a direct indicator of the Central Limit Effect and its changes in contour with increasing values of 2N show the approach to normality by the distribution of the mean $\overline{P-Q}$ of a sample of 2N observations drawn from a population of difference-scores, (P_i-Q_i)'s (or as henceforth designated, P-Q's). (It is shown later that <u>all</u> Z statistics investigated in this study can be regarded as transformations upon the mean of a single sample and that the robustness of the Z's is therefore merely a reflection of the Central Limit Effect. However it is only when samples are of equal size that the Central Limit Effect "applies" to a <u>conventional</u> population of difference-scores.)

All that is needed, therefore, to complete the picture is a knowledge of the shape of the difference-score population from which samples are "drawn". The true distribution of a difference-score can be obtained from the true distributions of the two variables from which the difference-score is calculated. For example, the true X-Y difference-score distribution is obtained by pairing every value of X in the X population (of which there are 120) with each of the numerically different values of Y to be found in the Y population (of which there are 97). The chance probability for each such pairing is the product of the point probabilities of the X value and Y value involved. This probability is calculated and so is the X-Y difference-score corresponding to that particular pairing. The probabilities for all pairings resulting in the same single difference-score are then summed, and this sum is the point probability of that particular difference-score in the true distribution of difference-scores. Using essentially this procedure, the true distributions of the difference-scores, X-X, X-Y, X-A, X-B, Y-A, Y-Y, and Y-B, were obtained by the IBM 7090 and printed out along with the mean and variance of each difference-score population and the sum of all difference-score point probabilities for each difference-score distribution. The latter were included as checks since it was known a priori that $\mu_{P-Q} = \mu_P - \mu_Q = 100-100 = 0$, $\sigma^2_{P-Q} = \sigma^2_P + \sigma^2_Q$, and $\sum$Prob (P-Q) = 1. The Y-Y and Y-B distributions were obtained because they are essentially normal distributions with the same mean, variance and area as X-X or X-Y, in the former case, and as X-A, X-B, or Y-A, in the latter. Thus they conveniently provided the data required to plot the appropriate normal distributions for comparison with the nonnormal X-X, X-Y, X-A, X-B, and Y-A distributions. (The Y-X, A-X, B-X, and A-Y difference-score distributions were not obtained since it is obvious from our previous considerations that they are the mirror images of the X-Y, X-A, X-B, and Y-A difference-score distributions respectively.) Figures showing the shapes of the various difference-score populations are presented later in order that they may accompany figures showing results appropriate to them. In these figures, the nonnormal distributions are represented by histograms, each difference-score being represented by a different histogram bar so that the "grainedness" of the histogram corresponds exactly to the discreteness of the population. Each histogram is accompanied by a normal distribution having the same mean, variance, and area and represented by a continuous curve which extends just far enough in each direction (i.e., ± 3.2905 standard deviations) to enclose the central 99.9% of the normal distribution.

RESULTS

I. Manner of Presentation of Results

At each value of N the computing machine printed out tables giving, for each statistic investigated, the empirical one-tailed cumulative probabilities of those values of the statistic having normal-theory one-tailed cumulative probabilities of .0005, .001, .005, .01, .025, .05, .10, .20, .30, .40, and .50, and this was done for both left and right tails. These tables were photographed, and the photographic reproductions of all such tables, i.e., the basic data, are presented in appendix IV. The present section of the report presents only derived data, in the form of graphs.

Let α be a standard significance level, such as .05, and let S_α be that value of the statistic S which under normal-theory (i.e., when the tested hypothesis and all assumptions are true) has a true cumulative probability of α . Finally, let ρ be the empirical cumulative probability of S_α in the present study. Then ρ is the empirical, and therefore chance-influenced, estimate of the <u>true</u> cumulative probability, $\bar{\rho}$, of S_α (when assumptions are violated but the tested hypothesis is true). Thus ρ is an estimate of the true significance level at which the test is actually being conducted. Consequently, except for the effects of chance upon ρ, the degree to which ρ corresponds with α serves as a direct indicator of the robustness of the test when conducted at a nominal significance level of α. If the test is ("quite") robust ρ should tend to equal α, and the ratio between ρ and α should tend to equal 1.00; to the extent that it departs from 1.00 to a degree unlikely to occur by chance, justification is lacking for the claim that the test is ("quite") robust. (The qualifications in parentheses are necessitated by the lack of any widely accepted objective definition of robustness.)

To illustrate, suppose one is interested in the robustness of TXX 2NN when the "first" and "second" samples are based upon 2N = 16 and N = 8 observations respectively and the test is a right-tailed test conducted at the nominal .05 level of significance. Consulting the table labeled N = 8 and reporting results only for T-statistics in appendix IV it is found that the entry for the row labeled TXX 2NN and the right hand column labeled 0500 is 97. This implies the following information. TXX 2NN at N = 8 is a two-sample t test with 16 + 8 - 2 =

22 degrees of freedom. The t value with 22 degrees of freedom which, under normal theory, corresponds to a significance level of .05 (i.e., the t value which forms the boundary of a right-tailed rejection region whose area is .05 if the tested hypothesis and all assumptions are true) is 1.7171. The right-tailed cumulative probability of the value 1.7171 in the empirical sampling distribution of TXX 2NN for N = 8 is .0097, and .0097 is therefore the value of ρ corresponding to an α of .05. Proceeding now to obtain the ratio ρ/α, this is found to be .194, from which it is apparent that the actual significance level is estimated to be approximately one-fifth of the alleged significance level.

The ratio ρ/α can be no smaller than zero (when $\rho = 0$) but can be as large as $\frac{1.00}{\alpha}$ (when $\rho = 1.00$). Thus, when $\alpha = .05$ the ratio ρ/α can be 1 less than 1.00, but could conceivably be 19 times greater than 1.00. This asymmetry about the value 1.00, which represents a condition of robustness, is undesirable because it gives greater weight to excesses of ρ over α than it does to deficits. This problem was partially solved by using the ratio ρ/α when $\rho < \alpha$ and its reciprocal, i.e., the ratio α/ρ, when $\rho > \alpha$, plotting the latter against a scale which is the inverse of the one used for the former and contiguous with it, the two scales having a common point at $\rho/\alpha = 1.00 = \alpha/\rho$. Thus, the lower portion of the ordinate scale, for values rising from 0 to 1, pertains to the ratio ρ/α. When ρ exceeds α the ratio ρ/α, had it been used, would have exceeded 1.00, and the greater the excess of ρ over α, the greater would have been the ratio ρ/α, and the farther above 1.00 would it have been plotted. However the same qualitative effect is obtained by plotting, instead, the ratio α/ρ upon the upper portion of the ordinate scale whose values descend from 1.00 to 0 as one goes physically higher upon the scale. Although the ρ/α and α/ρ scales are equal in length, the problem is only partially solved because while ρ can assume values small enough to make $\rho/\alpha = 0$, it cannot assume values large enough to make $\alpha/\rho = 0$. Thus points can be plotted at the very bottom of the graph, but not at the very top, and a measure of asymmetry remains. Furthermore, since the highest point which could possibly be plotted occurs when $\alpha/\rho = \alpha/1.00$, the "highest plottable point" varies with α. Despite its shortcomings, the procedure followed seemed to be the most meaningful way to present the data, and it was followed for all data presented in this section, both data relevant primarily to the Central Limit Effect and that relevant only to robustness. One final point should be mentioned.

The smaller α is the smaller ρ tends to become and the greater becomes the influence of chance upon ρ (i.e., the greater becomes the variability of ρ relative to the size of the true value which ρ estimates). For sampling distributions based on 10,000 samples, the expected frequency of statistic-values contributing to ρ, if the test statistic is robust, is approximately 500 when α = .05, 100 when α = .01 and only 10 when α = .001. The upper and lower tolerance limits about the expected, i.e., true, frequency of 500 are 543 and 457 at a .95 tolerance level, 556 and 444 at a .99 tolerance level, and 572 and 428 at a .999 tolerance level. About an expected frequency of 100, the tolerance limits are 120 and 80, 126 and 74, and 133 and 67 at the respective tolerance levels mentioned, and about an expected frequency of 10, the corresponding limits are 16 and 4, 18 and 2, and 20 and 0 at the respective tolerance levels listed. Thus, at least for robust test statistics, the curves depicting ρ/α or α/ρ can be expected to show very little jaggedness due to chance effects when α = .05 and to show only a moderate amount when α = .01, but to show a great deal when α = .001. Of course, this reasoning does not hold for tests which are quite nonrobust, and in these cases the curves may be and probably are highly reliable even when α = .001.

II. The Central Limit Effect as Influenced by Population Shape

Since the two-sample Z statistic, when based upon samples of equal size, is equivalent to the one-sample Z statistic based upon observations drawn from a population of difference scores, and since the one-sample Z statistic is simply a linear transformation upon the sample mean, the sampling distributions of the following two-sample Z statistics (in appendix code), ZXX 2N2N, ZXY 2N2N, ZYX 2N2N, ZXA 2N2N, ZAX 2N2N, ZXB 2N2N, ZBX 2N2N, ZYA 2N2N, and ZAY 2N2N, show the Central Limit Effect for the condition of sampling from the following, respective, difference-score populations, X-X, X-Y, Y-X, X-A, A-X, X-B, B-X, Y-A, and A-Y. The data for all nine situations are presented separately in appendix IV (as well as data for a tenth relevant situation, that of ZXX NN). However, in this section of the report data for the nine cases are combined (in pairs, except for X-X) to eliminate the redundancy of mirror-image true sampling distributions while using the additional "information" to increase precision of estimate. (This was discussed in a previous section.) The five cases resulting from these combinations are ZXX 2N2N, ZXY 2N2N, ZXA 2N2N, ZXB 2N2N and ZYA 2N2N representing samples from the respective difference-score populations, X-X, X-Y, X-A, X-B, Y-A. Thus the ρ corresponding to a "left-tailed" α of .01 for the "criss-cross combined" ZXY 2N2N (as graphed in the body of the report) is the average of the ρ corresponding to a "left-tailed" α of .01 for the "original" ZXY 2N2N and the ρ corresponding to a "right-tailed" α of .01 for ZYX 2N2N (as tabled in the appendix). Likewise the ρ corresponding to a "one-tailed" α of .01 for the "folded over" ZXX 2N2N (as graphed in the body of the report) is the average of the ρ corresponding to a "left-tailed" α of .01 for the "original" ZXX 2N2N with the ρ corresponding to a "right-tailed" α of .01 for the "original" ZXX 2N2N distribution (as tabled in the appendix). These data are reported here in terms of means, rather than Z's, since interest is centered upon the Central Limit Effect. However, the two are interchangeable in the context of the graphs presented, and a graph showing the Central Limit Effect for $\overline{X-Y}$ shows equally the Central Limit Effect for, and the robustness of, ZXY 2N2N.

The figures are presented sequentially in pairs, the first member of each pair showing the true (i.e., exact) shape of the population to which the second member, showing the Central Limit Effect, is relevant. The first pair of figures show the X population and the Central Limit Effect upon $\overline{X}$, i.e., the mean of samples drawn from it. The curves showing the Central Limit Effect are based upon sampling distributions of 50,000 values of $\overline{X}$ for each $N \leqq 1024$, and upon 10,000 values at higher N's. Thus the first Central Limit graph is unique in that it is derived from a Z statistic which was "originally" a one-sample Z statistic (otherwise stated, the sampled population is a population of absolute scores rather than difference scores), and the graph is also unique in that its precision of estimate is greater, at $N \leqq 1024$, than is the precision of those which follow. The remaining Central Limit graphs are relevant to sampling from populations of difference-scores and are either based upon full sampling distributions each of which contains 20,000 values of a difference-score mean, under a given condition of sample size, or upon a half-distribution containing 10,000 values.

The figures showing the sampled populations are distorted at very small ordinate values. If all abscissa values were represented by ordinate values of the proper scale, the histogram bars representing the very low probability abscissa values would not be high enough to register on a photograph and to

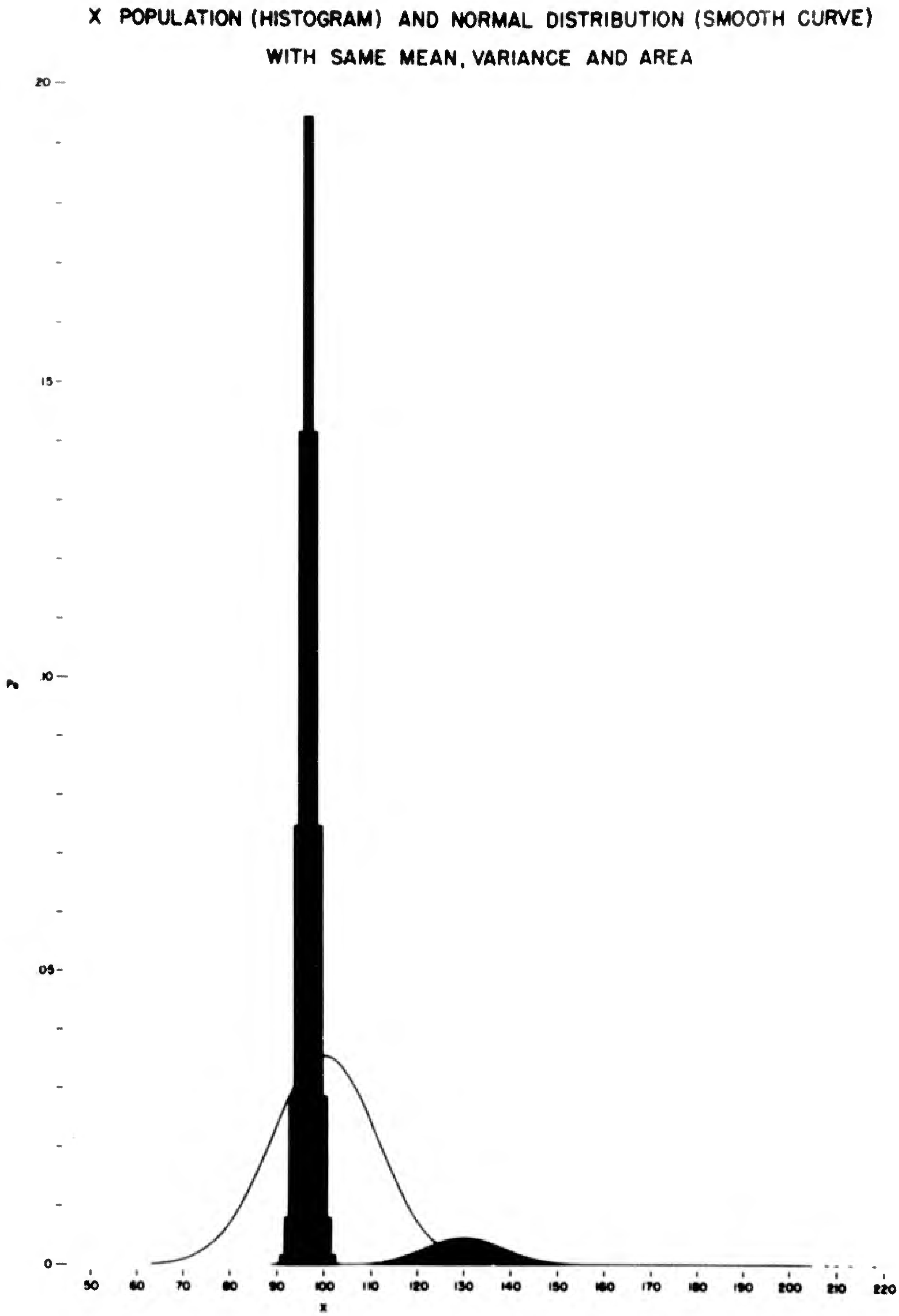

Figure 2. The X Population

CENTRAL LIMIT EFFECT UPON MEANS OF N OBSERVATIONS DRAWN FROM THE X POPULATION

RATIO BETWEEN EMPIRICAL, P, AND NORMAL-THEORY, α, ONE-TAILED CUMULATIVE PROBABILITIES FOR $\bar{X}$'S HAVING NORMAL-THEORY CUMULATIVE PROBABILITY OF α

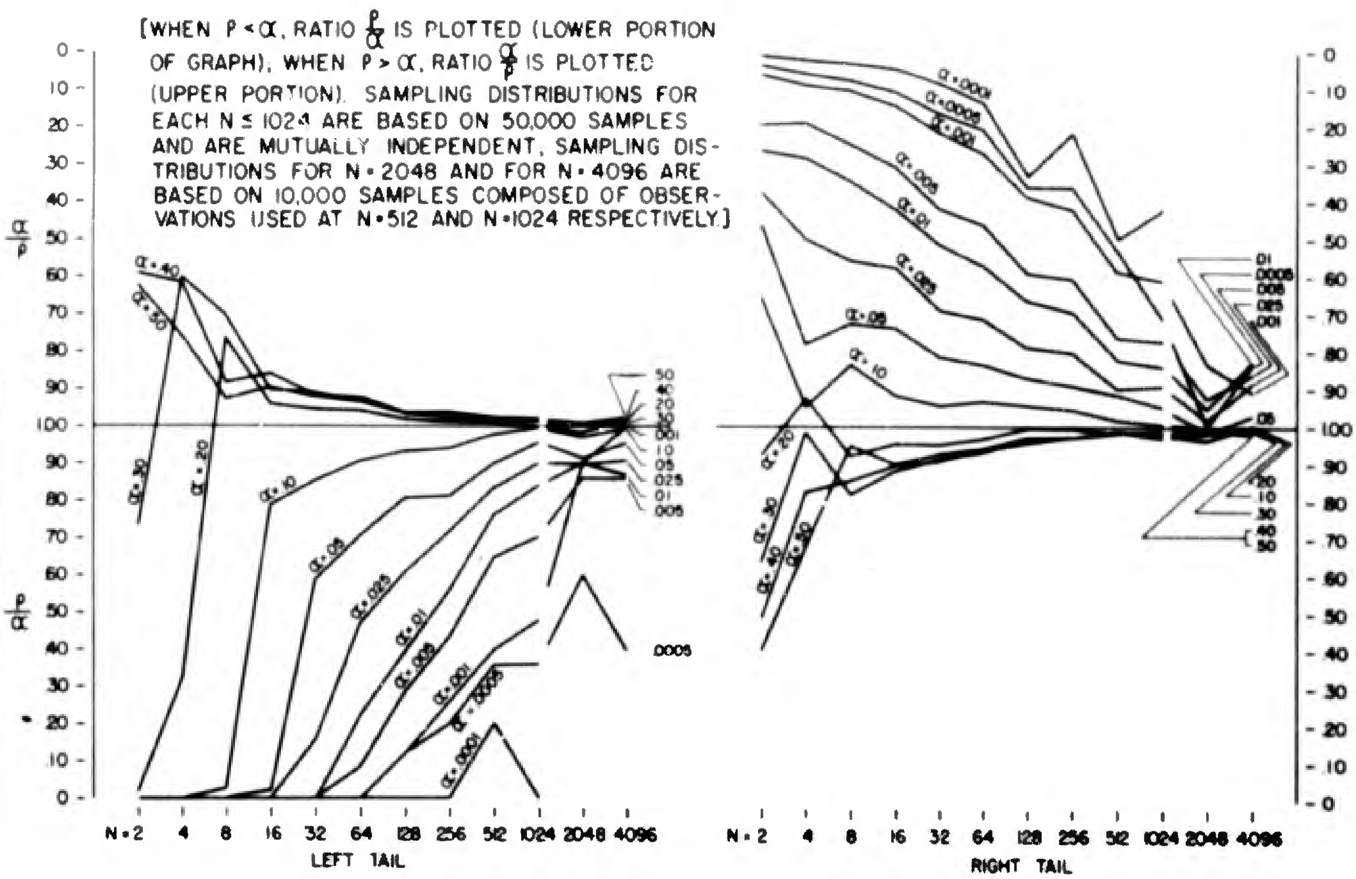

Figure 3. Central Limit Effect upon $\bar{X}$

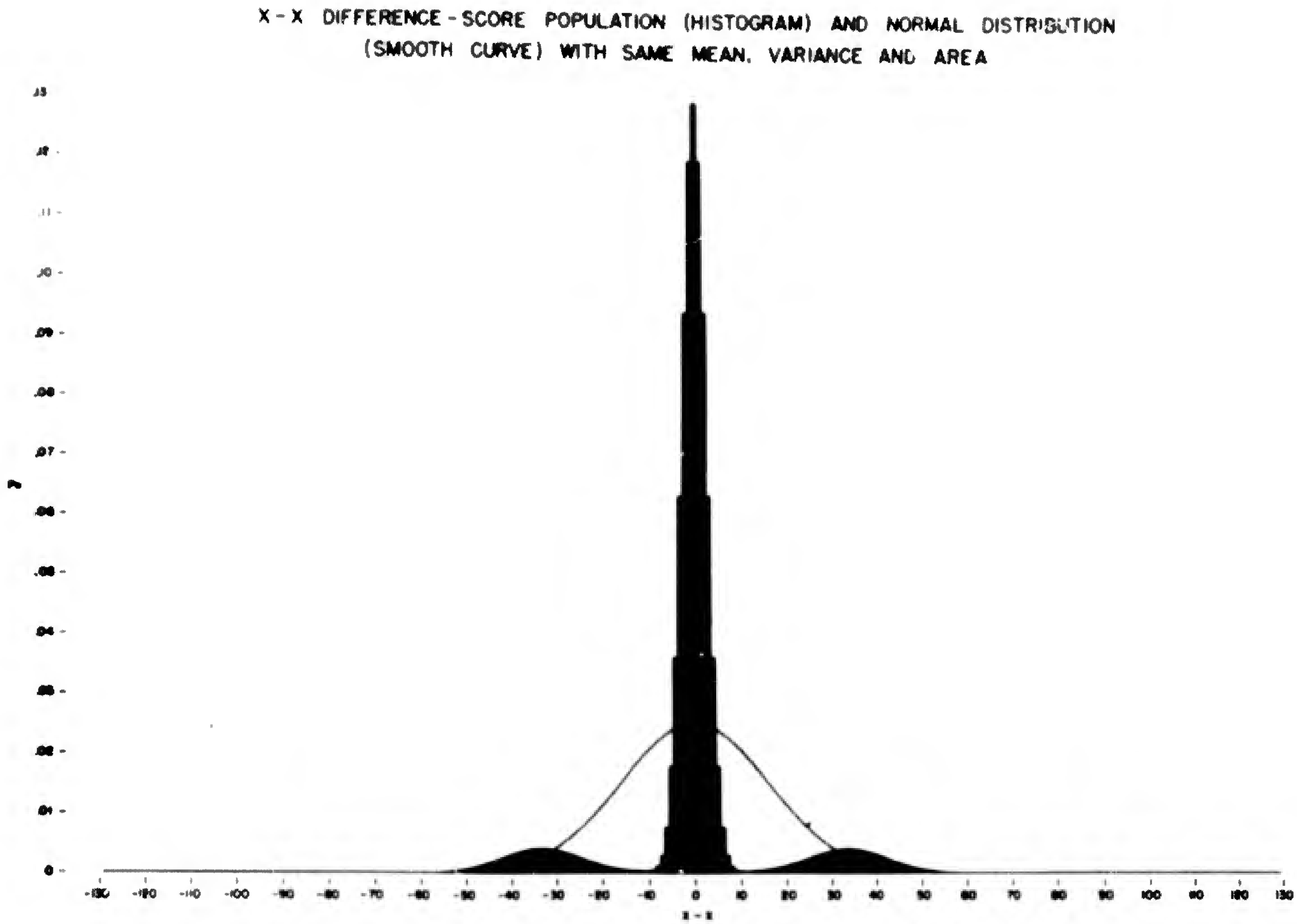

Figure 4. The X-X Difference-Score Population

CENTRAL LIMIT EFFECT UPON MEANS OF SAMPLES OF 2N OBSERVATIONS DRAWN FROM THE X-X DIFFERENCE-SCORE POPULATION

(ρ = EMPIRICAL CUMULATIVE PROBABILITY OF THAT VALUE OF $\overline{X-X}$ HAVING NORMAL-THEORY CUMULATIVE PROBABILITY OF α)

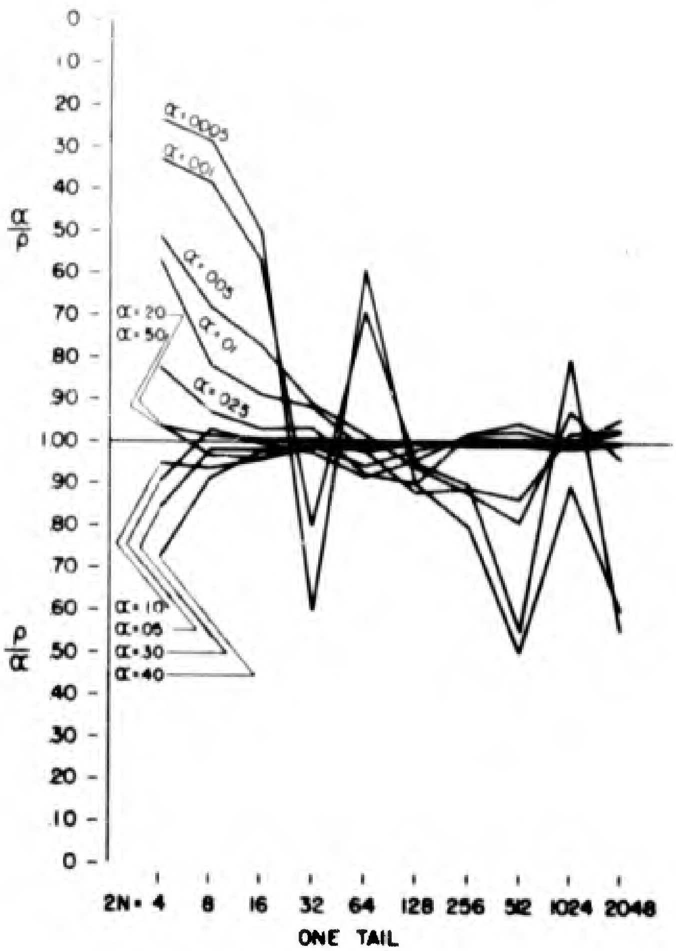

Figure 5. Central Limit Effect upon $\overline{X-X}$

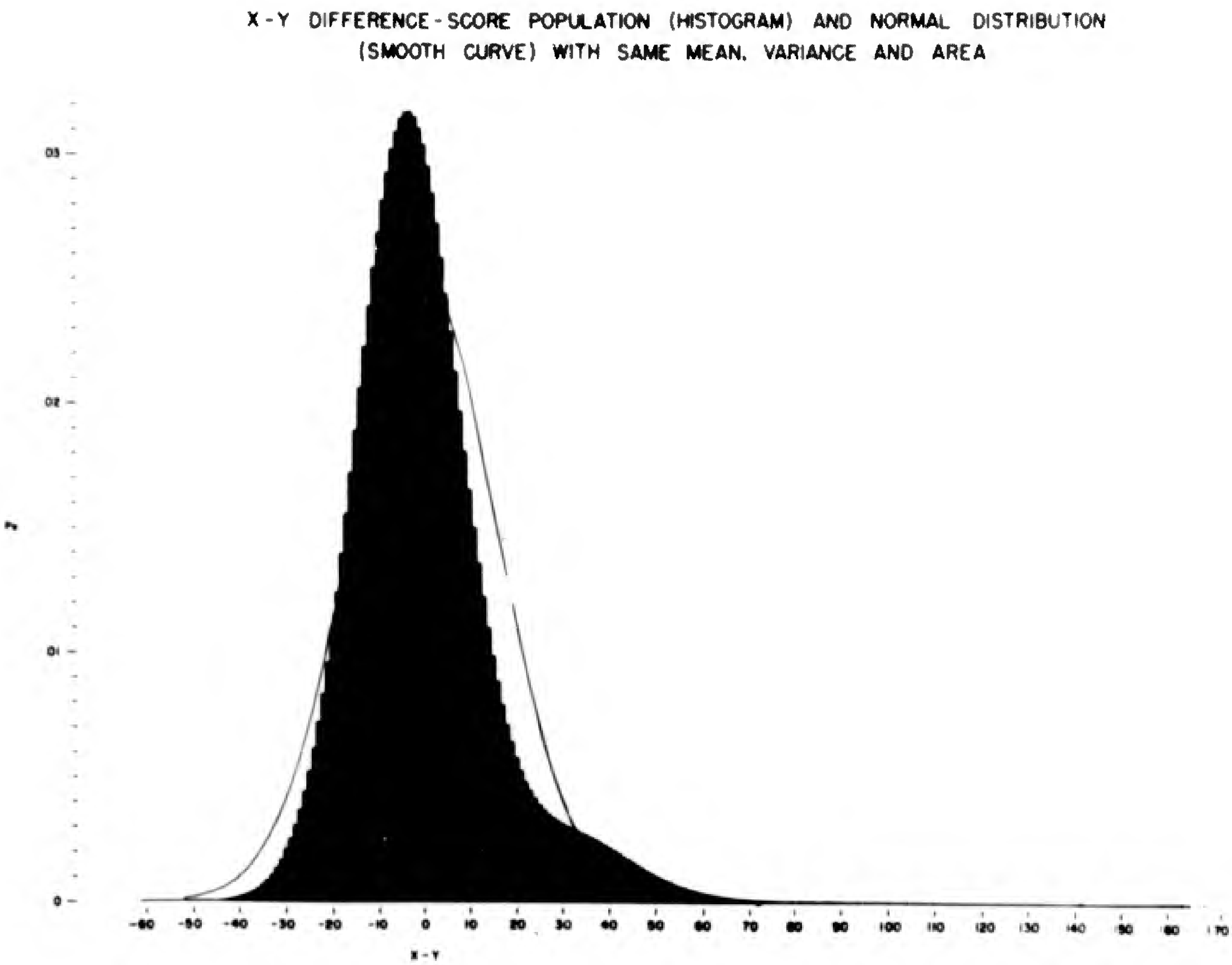

Figure 6. The X-Y Difference-Score Population

CENTRAL LIMIT EFFECT UPON MEANS OF SAMPLES OF 2N OBSERVATIONS DRAWN FROM THE X-Y DIFFERENCE-SCORE POPULATION

(ρ = EMPIRICAL CUMULATIVE PROBABILITY OF THAT VALUE OF $\overline{X\text{-}Y}$ HAVING NORMAL-THEORY CUMULATIVE PROBABILITY OF α)

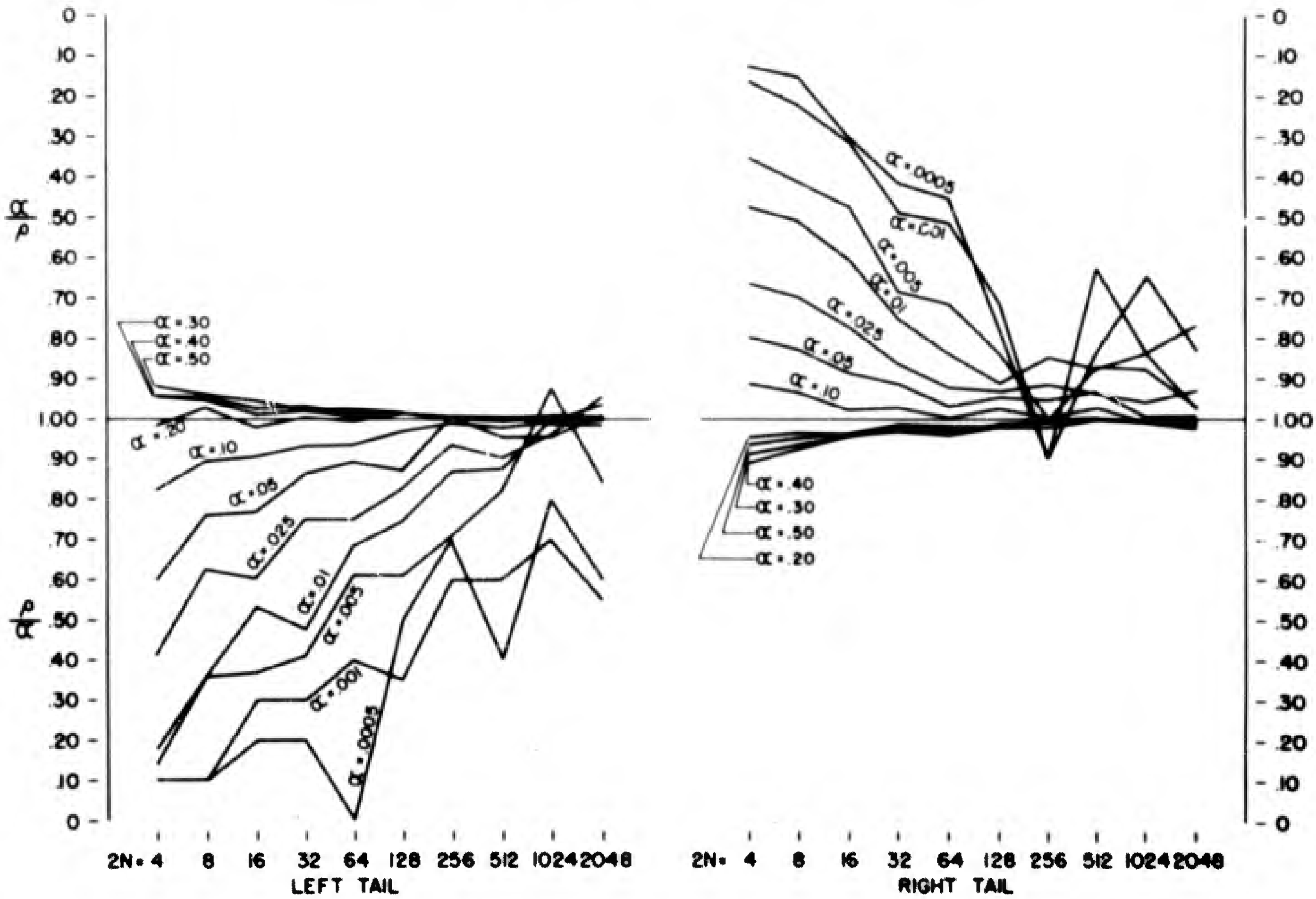

Figure 7. Central Limit Effect upon $\overline{X\text{-}Y}$

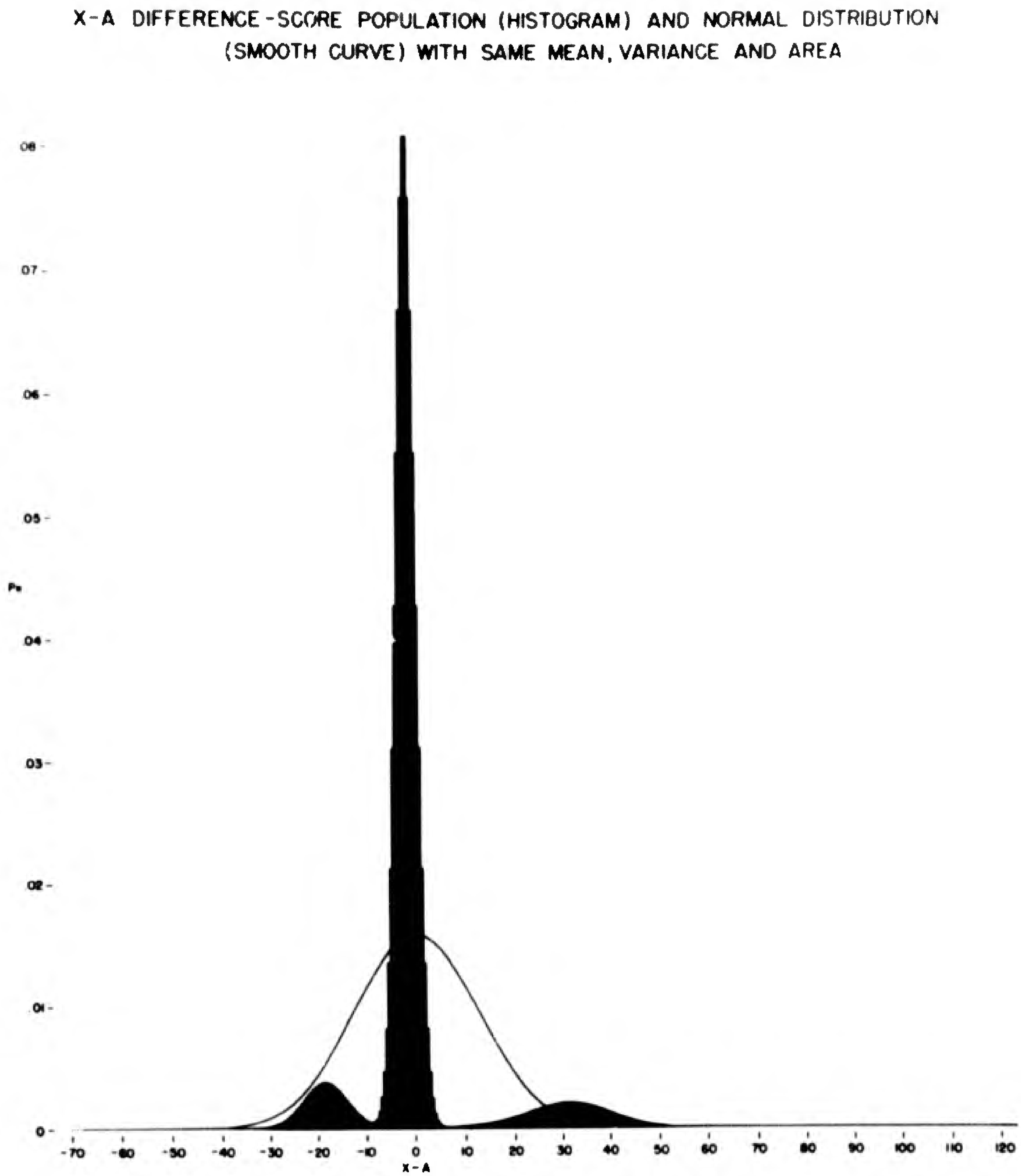

Figure 8. The X-A Difference-Score Population

CENTRAL LIMIT EFFECT UPON MEANS OF SAMPLES OF 2N OBSERVATIONS DRAWN FROM THE X-A DIFFERENCE-SCORE POPULATION

(ρ=EMPIRICAL CUMULATIVE PROBABILITY OF THAT VALUE OF $\overline{X\text{-}A}$ HAVING NORMAL-THEORY CUMULATIVE PROBABILITY OF α)

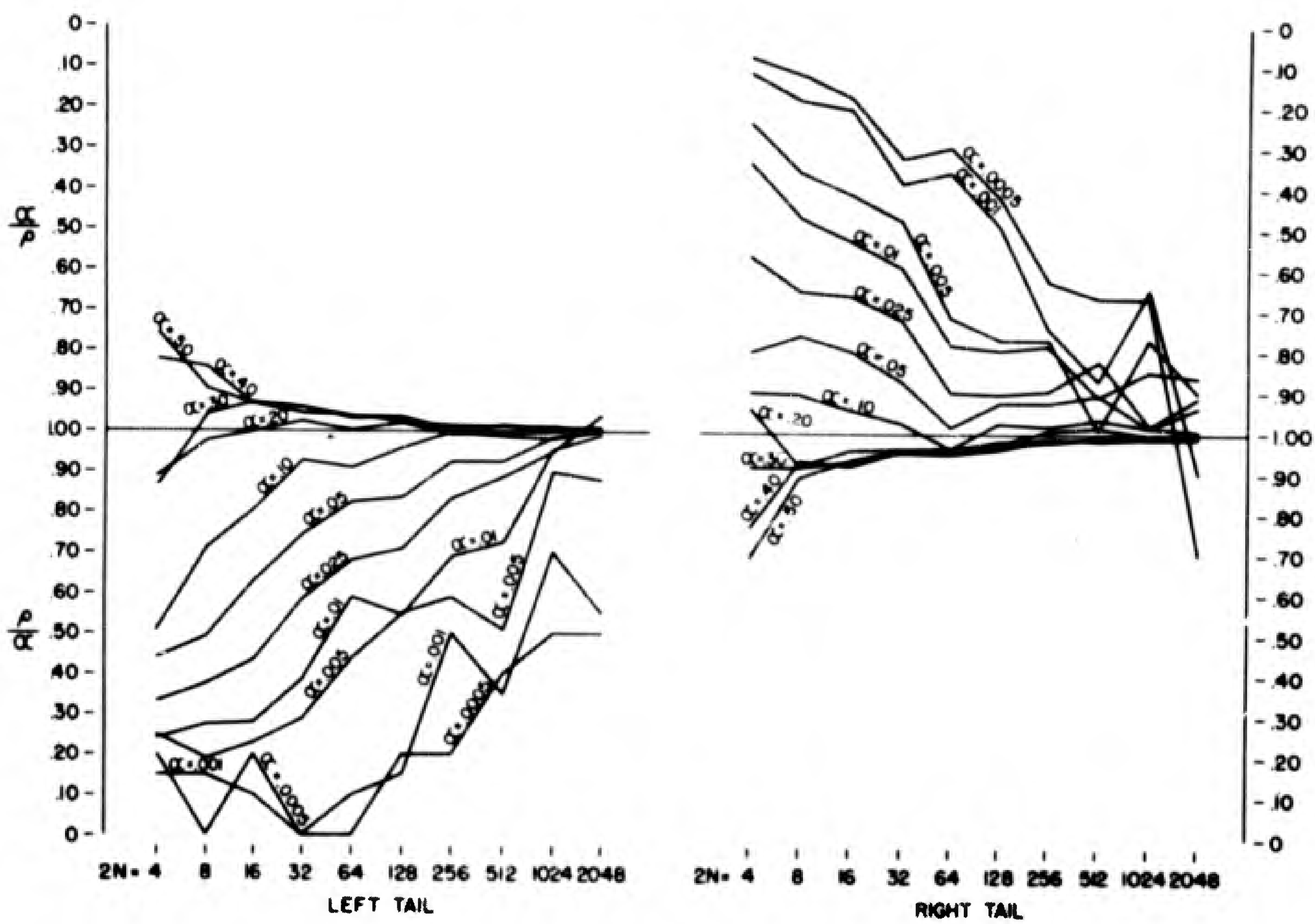

Figure 9. Central Limit Effect upon $\overline{X\text{-}A}$

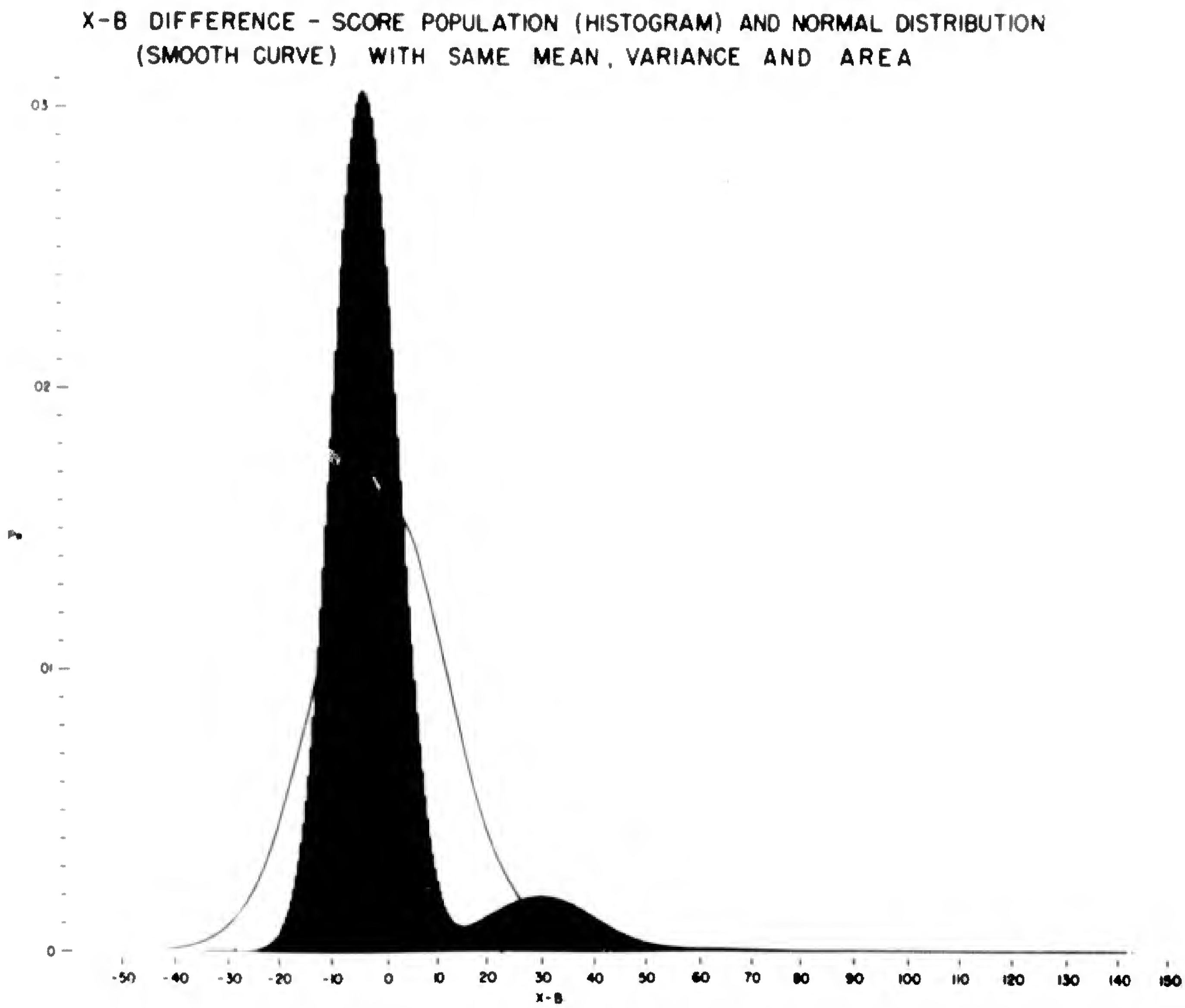

Figure 10. The X-B Difference-Score Population

CENTRAL LIMIT EFFECT UPON MEANS OF SAMPLES OF 2N OBSERVATIONS DRAWN FROM THE X-B DIFFERENCE-SCORE POPULATION

(ρ = EMPIRICAL CUMULATIVE PROBABILITY OF THAT VALUE OF $\overline{X-B}$ HAVING NORMAL-THEORY CUMULATIVE PROBABILITY OF α)

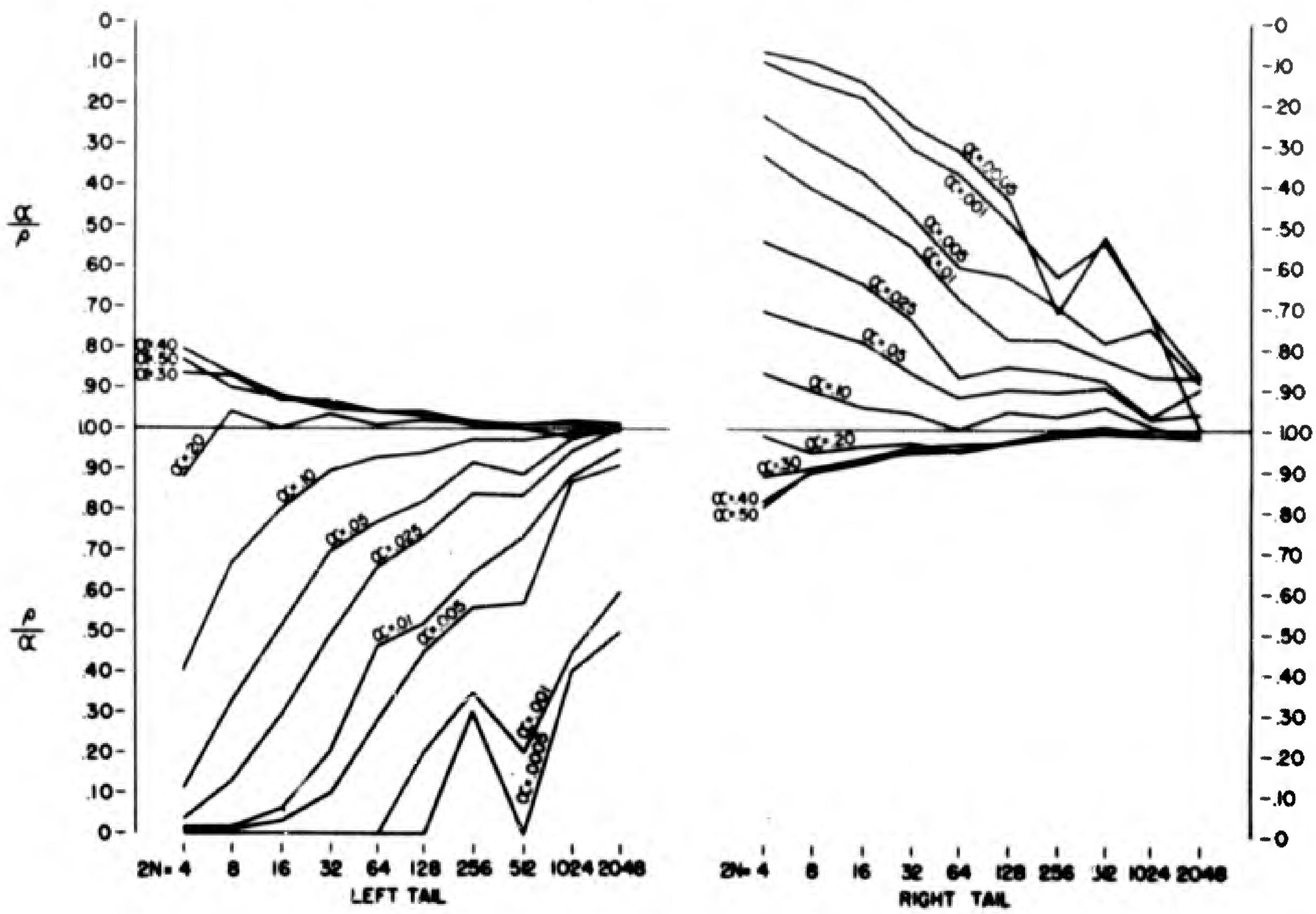

Figure 11. Central Limit Effect upon $\overline{X-B}$

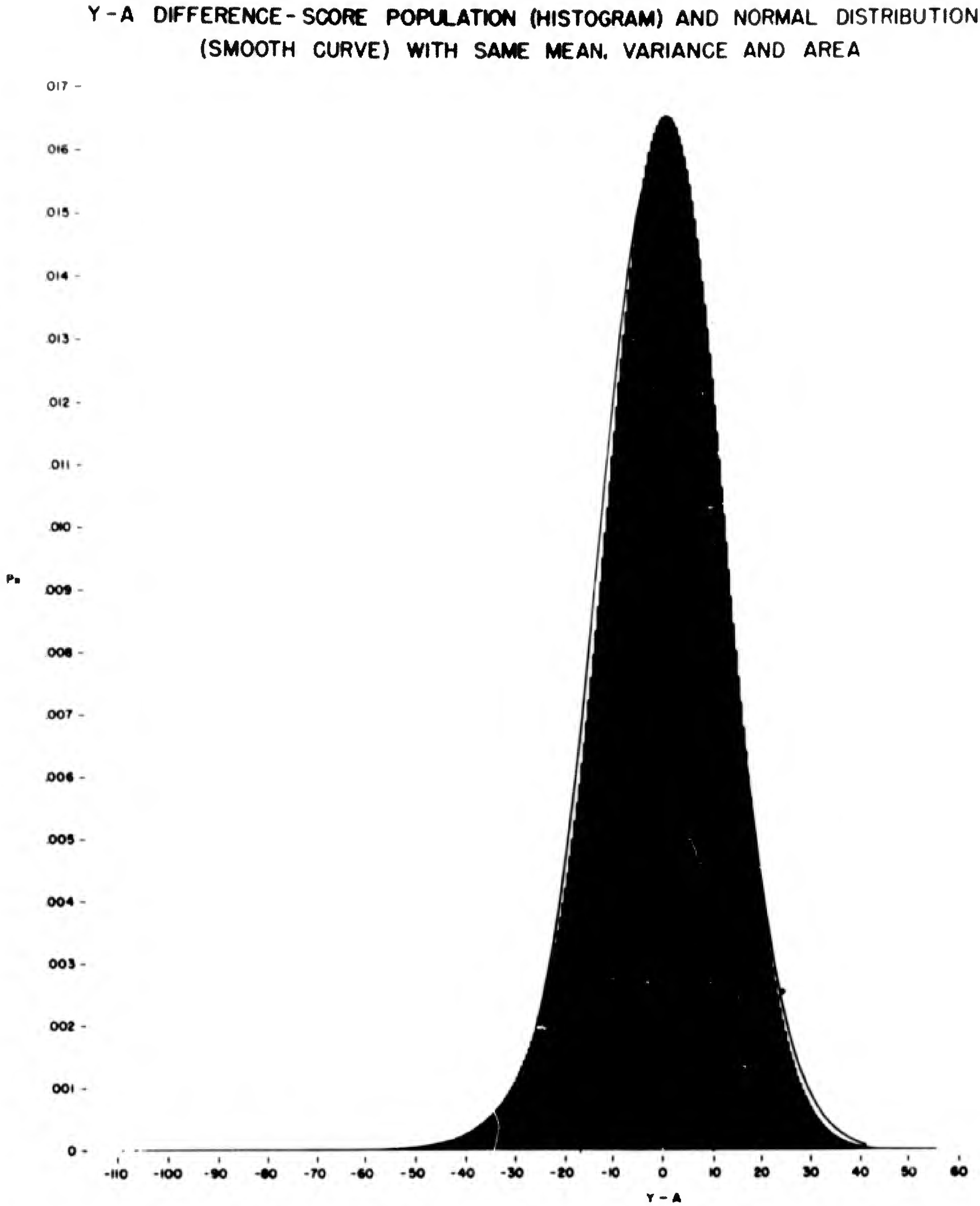

Figure 12. The Y-A Difference-Score Population

CENTRAL LIMIT EFFECT UPON MEANS OF SAMPLES OF 2N OBSERVATIONS DRAWN FROM THE Y-A DIFFERENCE-SCORE POPULATION

(ρ = EMPIRICAL CUMULATIVE PROBABILITY OF THAT VALUE OF $\overline{Y\text{-}A}$ HAVING NORMAL-THEORY CUMULATIVE PROBABILITY OF α)

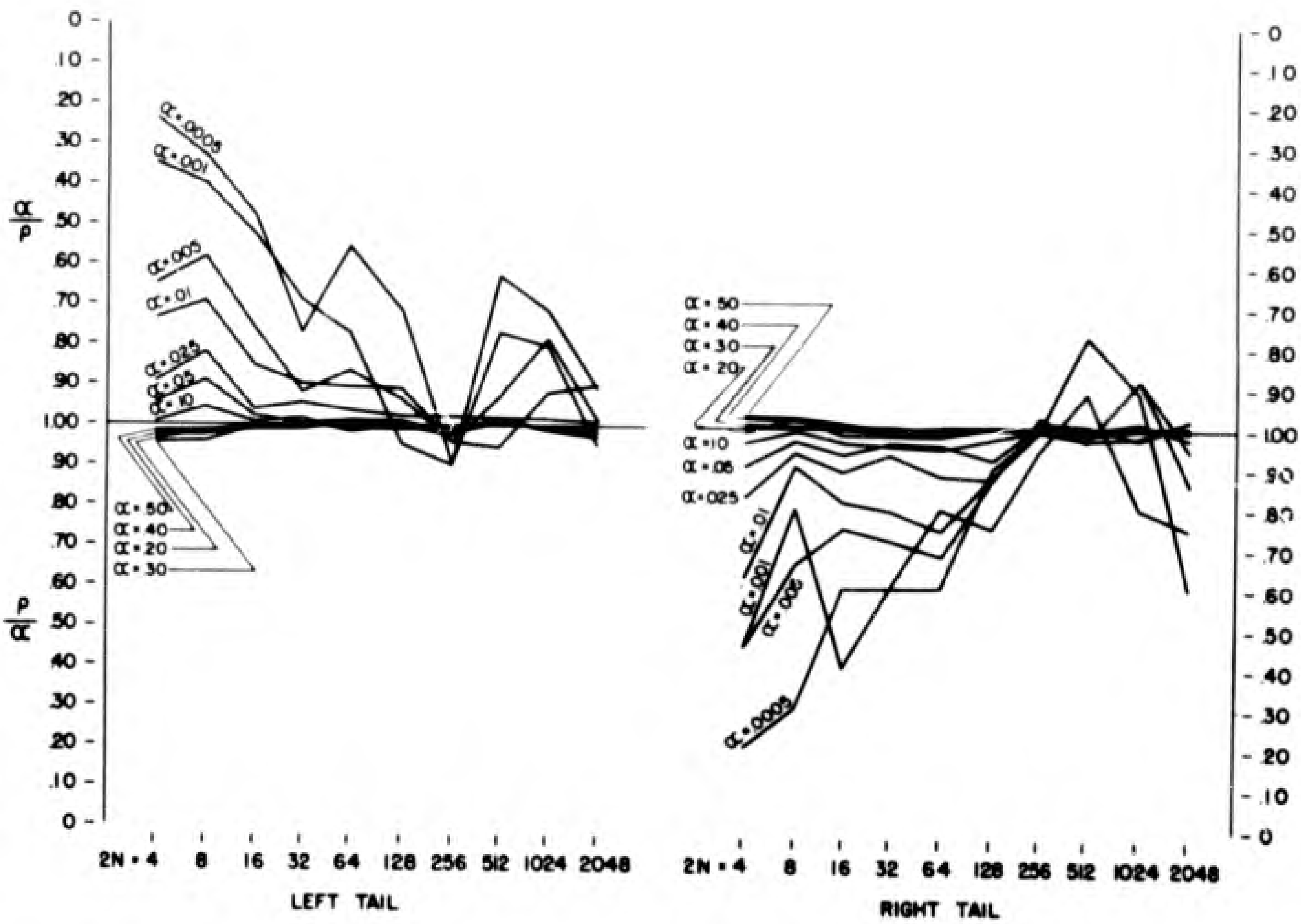

Figure 13. Central Limit Effect upon $\overline{Y\text{-}A}$

survive the duplicating processes involved in publication. Therefore all probabilities below the ordinate value corresponding to the line width produced by a certain drafting pen are represented by that line width and that ordinate. The result is that the thinnest tails of the figures are exaggerations of the true tails.

III. Robustness of the One-Sample t Test

Figure 14 shows the robustness of the one-sample t test when samples are drawn from the X population. It is analogous to figure 3 which, in effect, shows the robustness of the one-sample Z test under the same condition. Both figures are based on sampling distributions of 50,000 values for each $N \leqq 1024$, and as a result are, for the region in which $N \leqq 1024$, considerably more accurate at a given value of α, than the other graphs presented in this report.

IV. Robustness of Multi-Sample Tests

As was done in the case of graphs showing the Central Limit Effect, the redundancy of mirror-image true sampling distributions will be eliminated, but the precision of estimate will be increased, by combining data relevant to true sampling distributions, or portions thereof, which are mirror images of each other. This was discussed in an earlier section. There are four situations, illustrated as follows: (a) the ρ corresponding to right-tailed α of .05 for the combined, i.e., graphed, TXY 2N2N is the average of the ρ corresponding to a right-tailed α of .05 for the original (i.e., tabled in the appendix), TXY 2N2N and the ρ corresponding to a left-tailed α of .05 for the original TYX 2N2N, (b) the ρ corresponding to a two-tailed α of .05 for the combined TXY 2N2N is one half the sum of the four ρ's corresponding to one-tailed α's of .025 for the left and right tails of both TXY 2N2N and TYX 2N2N, i.e., it is the average of the two ρ's corresponding to two-tailed α's of .05 for TXY 2N2N and TYX 2N2N, (c) the ρ corresponding to a one-tailed α of .01 for the "folded over" TXX 2N2N is the average of the ρ corresponding to a left-tailed α of .01 for the original TXX 2N2N and the ρ corresponding to a right-tailed α of .01 for the original TXX 2N2N, (d) the ρ corresponding to a one-tailed α of .001 for the combined and folded over TYB 2N2N is the average of the ρ's corresponding to: a left-tailed α of .001 for the original TYB 2N2N, a right-tailed α of .001 for the original TYB 2N2N, a left-tailed α of .001 for the original TBY 2N2N, and a right-tailed α of .001 for the original TBY 2N2N.

ROBUSTNESS OF THE ONE-SAMPLE t TEST BASED ON N OBSERVATIONS DRAWN FROM THE X POPULATION

RATIO BETWEEN EMPIRICAL, ρ, AND NORMAL-THEORY, α, ONE-TAILED CUMULATIVE PROBABILITIES FOR t'S HAVING NORMAL-THEORY CUMULATIVE PROBABILITY OF α

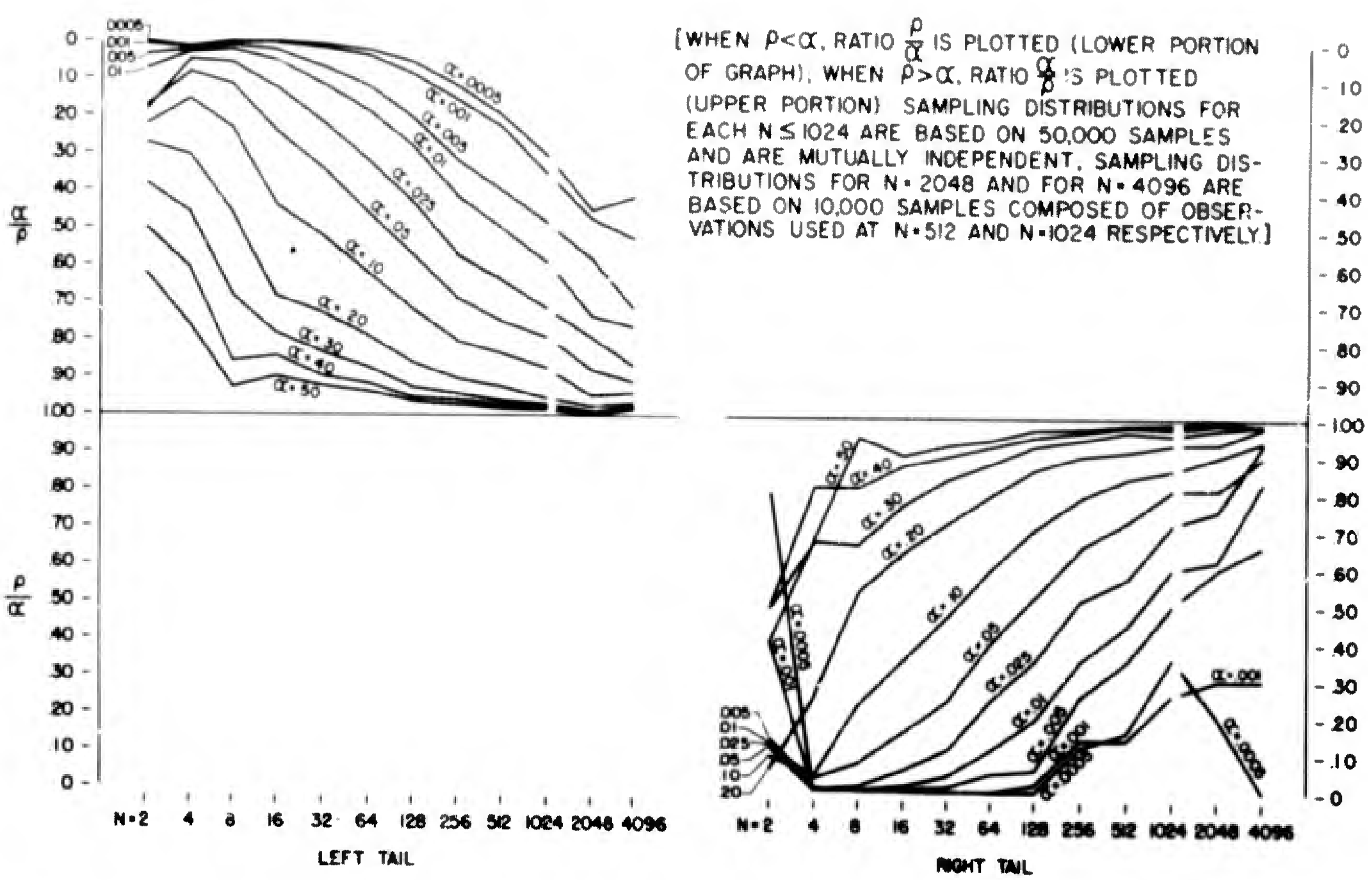

Figure 14. Robustness of TX

In the case of all statistics except TYB and TBY, separate graphs are presented to show robustness for left-tailed, two-tailed and right-tailed tests, i.e., for the cases of left-, two- or right-tailed rejection regions, and these graphs are restricted to three standard α levels, .05, .01, and .001. However, in the case of the TYB and TBY statistics, all of whose true sampling distributions are symmetrical, a single graph is presented for the case of a one-tailed test and six α levels are treated, .05, .025, .01, .005, .001, and .0005. The curves for "one-tailed" α's of .05, .01, and .001 apply to these same α levels for either a left-tailed or right-tailed test, i.e., apply to either type of one-tailed test. The curves for "one-tailed" α's of .025, .005, and .0005 apply to either left-tailed or right-tailed tests at those levels, but also apply to two-tailed tests at the .05, .01, and .001 levels respectively. Thus the single graph for TYB and TBY statistics presents the same type of information as that carried by all three graphs for the other Z and T statistics.

Since only the upper tail of the F statistic is used, ordinarily, as the rejection region for statistical tests, the F graphs presented in this section apply only to a "right-tailed" test. However when based upon two samples the right-tailed F test is equivalent to a two-tailed t test, i.e., the square of the value of t with n degrees of freedom corresponding to two-tailed significance level α, or one-tailed significance level $\alpha/2$, equals the value of F with 1 and n degrees of freedom corresponding to right-tailed significance level of α. Therefore (a) the right-tailed F test performs the same function as a two-tailed t test in the sense that it rejects for extreme discrepancies between means irrespective of the direction in which the extreme deviation (or deviations) occur, (b) graphs showing the robustness of the two-tailed t test are equivalently graphs showing the robustness of the F test based upon <u>two</u> samples, rather than the three or four formally investigated in this study, at the α value listed.

Just as the two-sample Z statistic, based upon equal sized samples, can be regarded as a one-sample Z statistic based upon observations drawn from a population of difference scores, so can the one-sample Z statistic be regarded as a two-sample Z statistic based upon samples one of which is the same as that upon which the one-sample Z statistic is based and the other of which is drawn from the same population but is infinite in size. Consider the formula for the two-sample Z statistic

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2_1}{N_1} + \frac{\sigma^2_2}{N_2}}}$$

If the two samples are drawn from the same population, $\mu_1 = \mu_2 = \mu$ and $\sigma^2_1 = \sigma^2_2 = \sigma^2$ so the formula can be written

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu - \mu)}{\sqrt{\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}}}$$

$$= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma^2}{N_1} + \frac{\sigma^2}{N_2}}}$$

Now let the second sample be infinite in size. Then $\sigma^2/N_2 = \sigma^2/\infty = 0$ and, by the Law of Large Numbers, $\bar{X}_2 = \mu$. Making these substitutions, we obtain

$$Z = \frac{\bar{X}_1 - \mu}{\sqrt{\frac{\sigma^2}{N_1}}}.$$

which is simply the formula for the one-sample Z based upon N_1 observations. Therefore the graphs in the section on the Central Limit Theorem which show the robustness of the "one-sample" Z statistics listed below on the left are just as validly graphs showing the robustness of the two-sample Z statistics listed on the right beside the one-sample statistics to which they are equivalent.

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{N}}} \qquad\qquad Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma^2_1}{N} + \frac{\sigma^2_2}{\infty}}}$$

$$Z = \frac{\overline{X\text{-}X} - \mu_{X\text{-}X}}{\sqrt{\frac{\sigma^2_{X\text{-}X}}{2N}}} \qquad Z = \frac{(\overline{X\text{-}X})_1 - (\overline{X\text{-}X})_2}{\sqrt{\frac{\sigma^2_{X\text{-}X}}{2N} + \frac{\sigma^2_{X\text{-}X}}{\infty}}}$$

$$Z = \frac{\overline{X\text{-}A} - \mu_{X\text{-}A}}{\sqrt{\frac{\sigma^2_{X\text{-}A}}{2N}}} \qquad Z = \frac{(\overline{X\text{-}A})_1 - (\overline{X\text{-}A})_2}{\sqrt{\frac{\sigma^2_{X\text{-}A}}{2N} + \frac{\sigma^2_{X\text{-}A}}{\infty}}}$$

$$Z = \frac{\overline{X\text{-}B} - \mu_{X\text{-}B}}{\sqrt{\frac{\sigma^2_{X\text{-}B}}{2N}}} \qquad Z = \frac{(\overline{X\text{-}B})_1 - (\overline{X\text{-}B})_2}{\sqrt{\frac{\sigma^2_{X\text{-}B}}{2N} + \frac{\sigma^2_{X\text{-}B}}{\infty}}}$$

$$Z = \frac{\overline{Y\text{-}A} - \mu_{Y\text{-}A}}{\sqrt{\frac{\sigma^2_{Y\text{-}A}}{2N}}} \qquad Z = \frac{(\overline{Y\text{-}A})_1 - (\overline{Y\text{-}A})_2}{\sqrt{\frac{\sigma^2_{Y\text{-}A}}{2N} + \frac{\sigma^2_{Y\text{-}A}}{\infty}}}$$

Thus the graphs showing the Central Limit Effect are relevant to the robustness of two-sample Z tests and extend beyond the cases examined for the latter in the sense that they show robustness at the maximal inequality of sample sizes (to the extent that maximal inequality is obtained when one sample is of finite size, the other infinite). One should not assume, however, that maximal inequality of sample sizes necessitates or implies maximum nonrobustness, and, to a considerable degree, the equivalence discussed is of academic interest since infinite sized samples are never obtained.

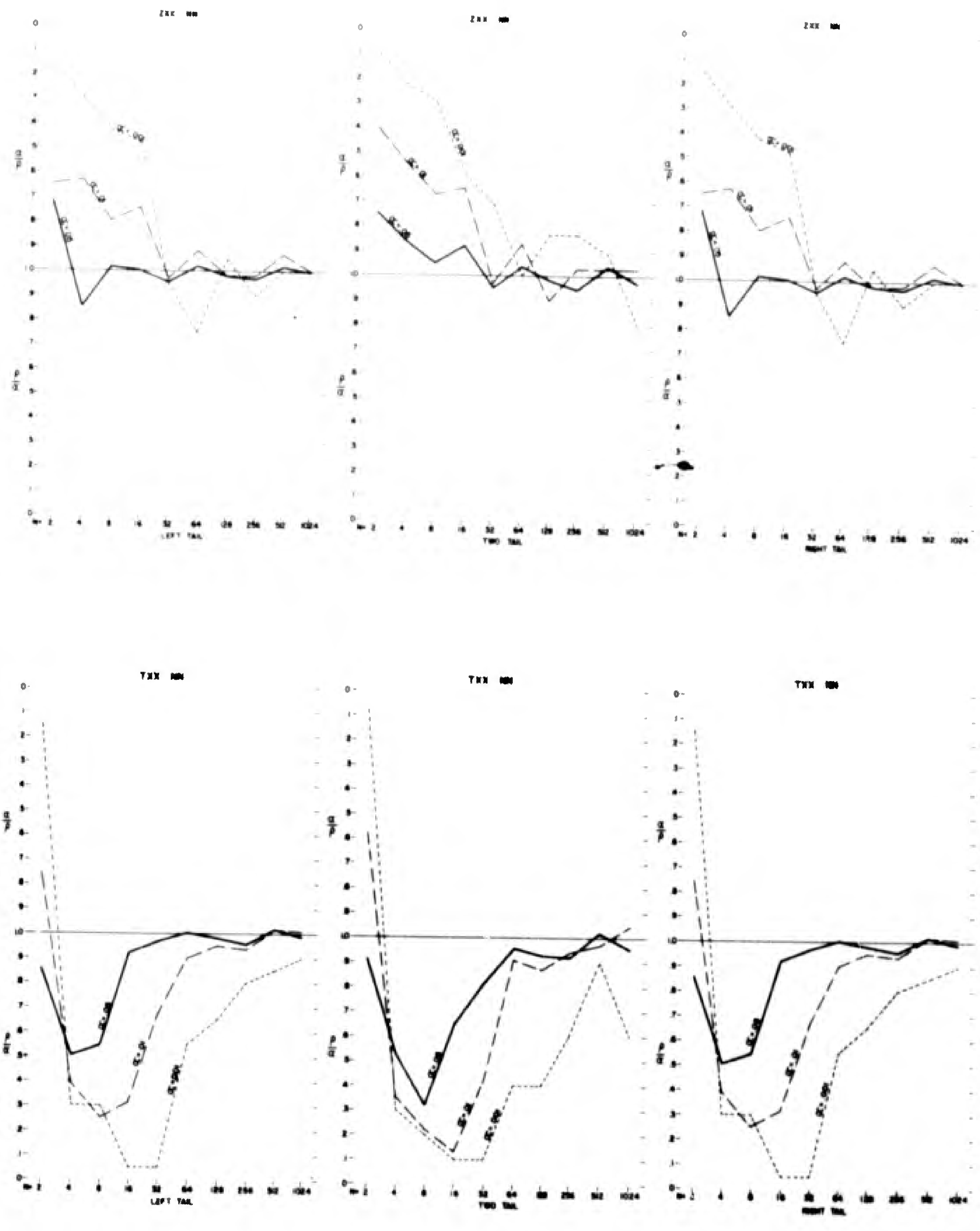

Figure 15. Robustness of ZXX NN and TXX NN

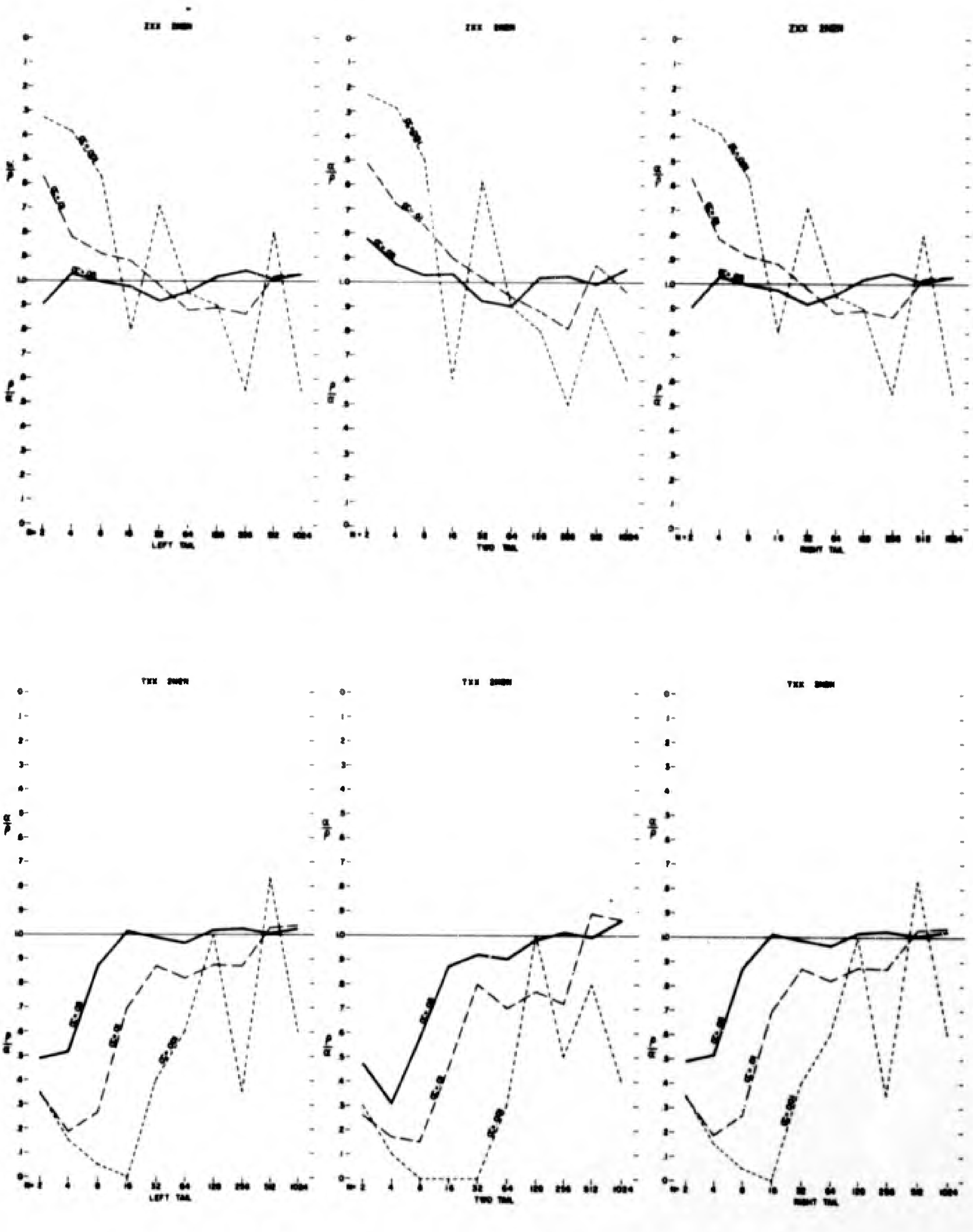

Figure 16. Robustness of ZXX 2N2N and TXX 2N2N

BLANK PAGE

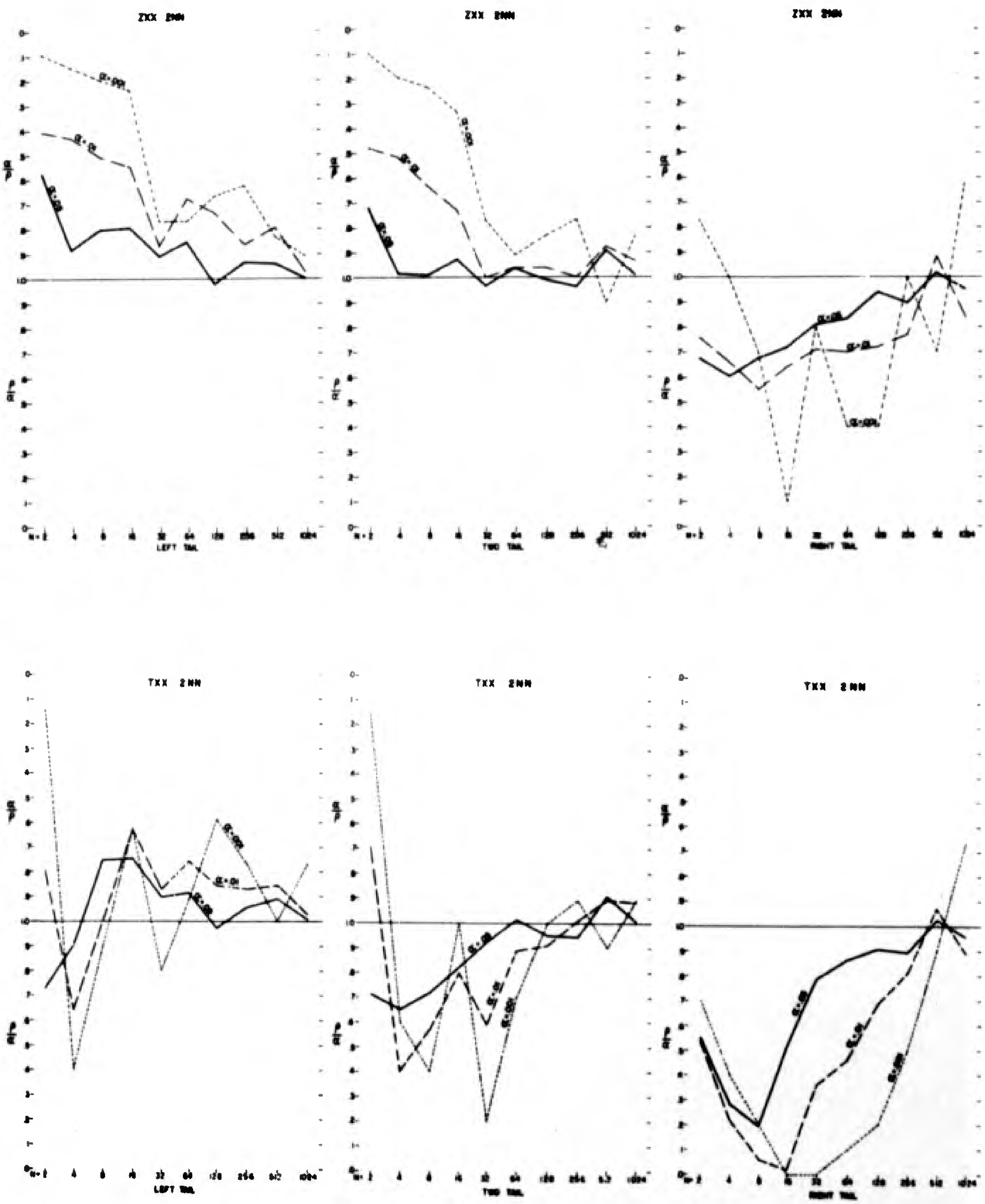

Figure 17. Robustness of ZXX 2NN and TXX 2NN

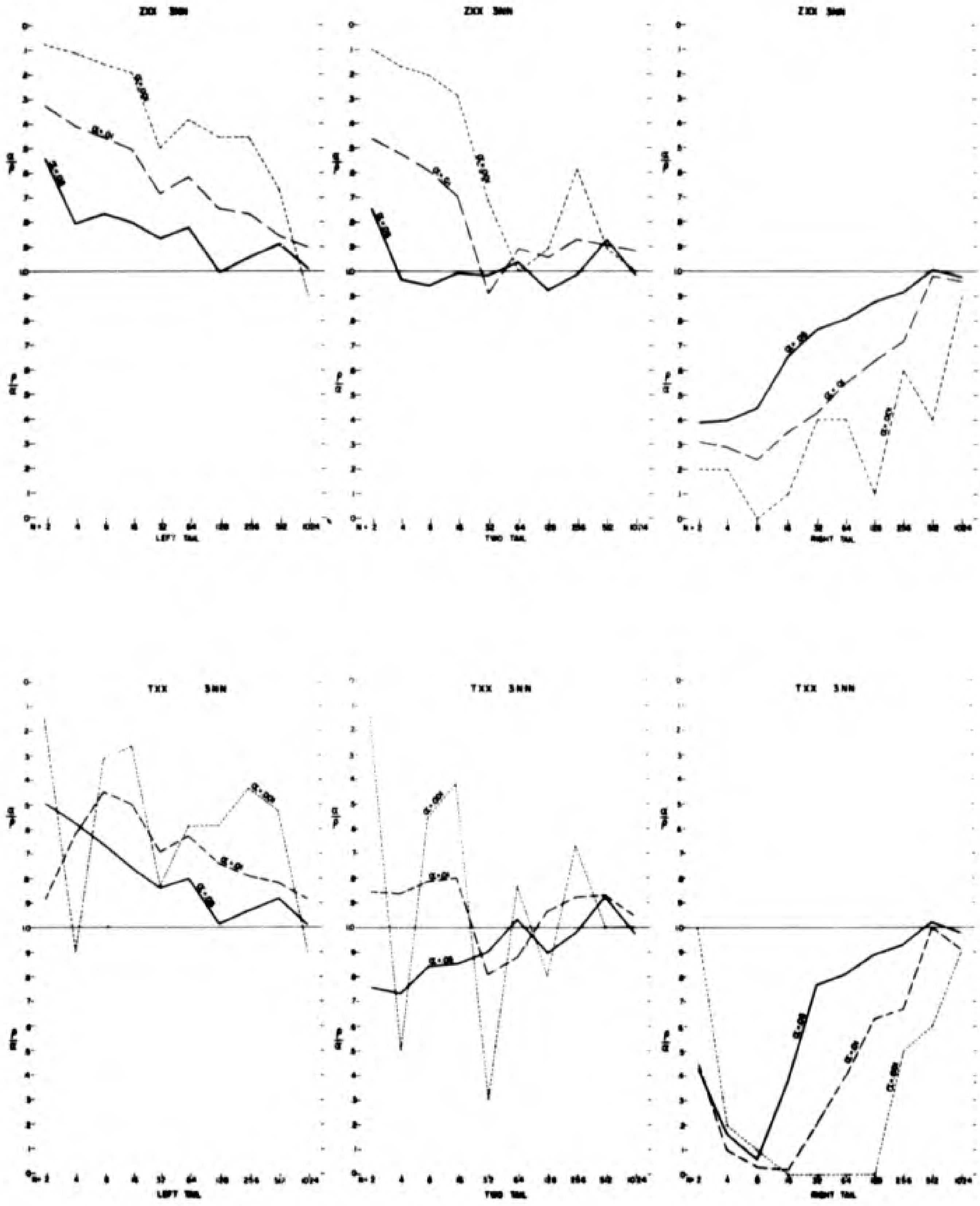

Figure 18. Robustness of ZXX 3NN and TXX 3NN

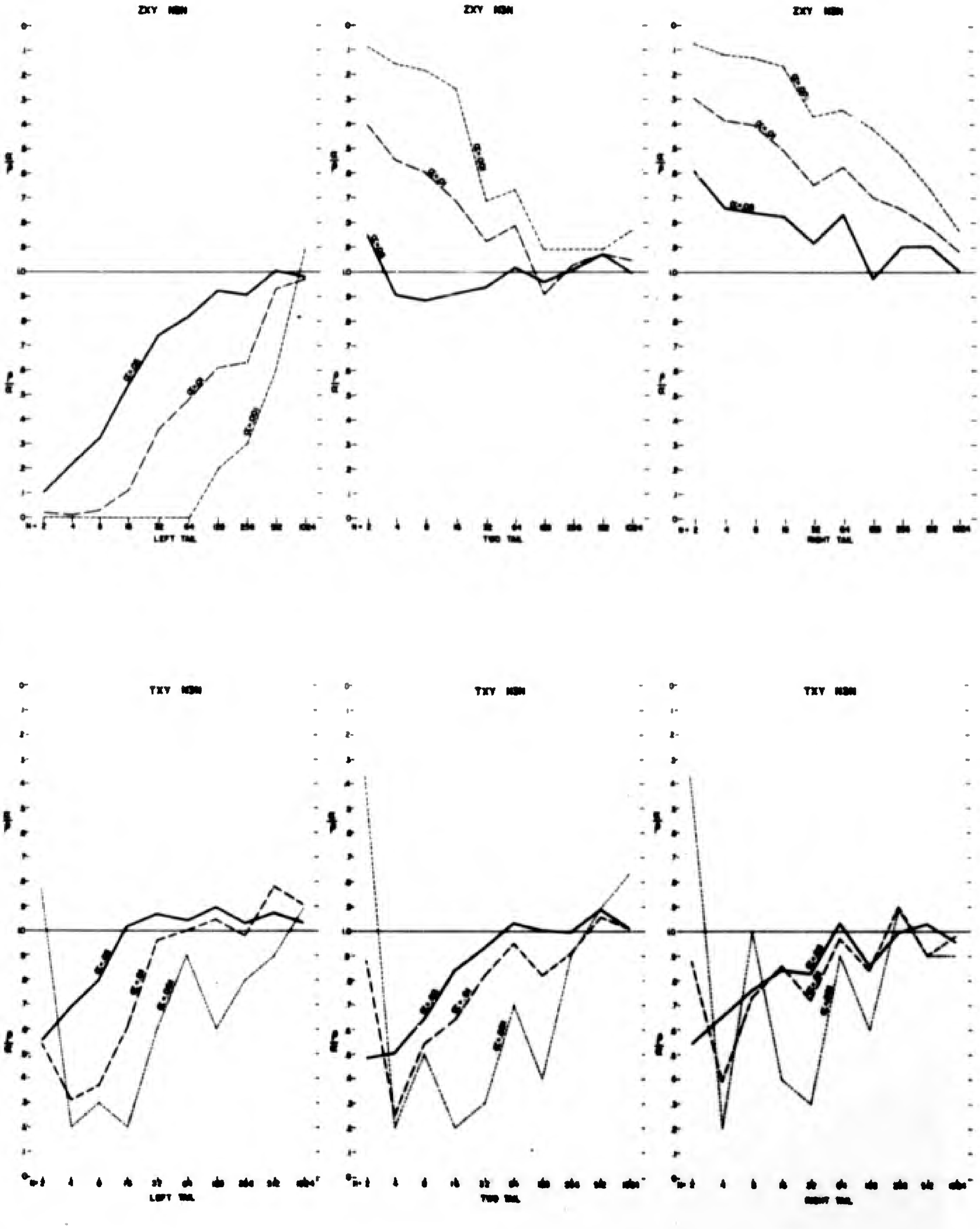

Figure 19. Robustness of ZXY N3N and TXY N3N

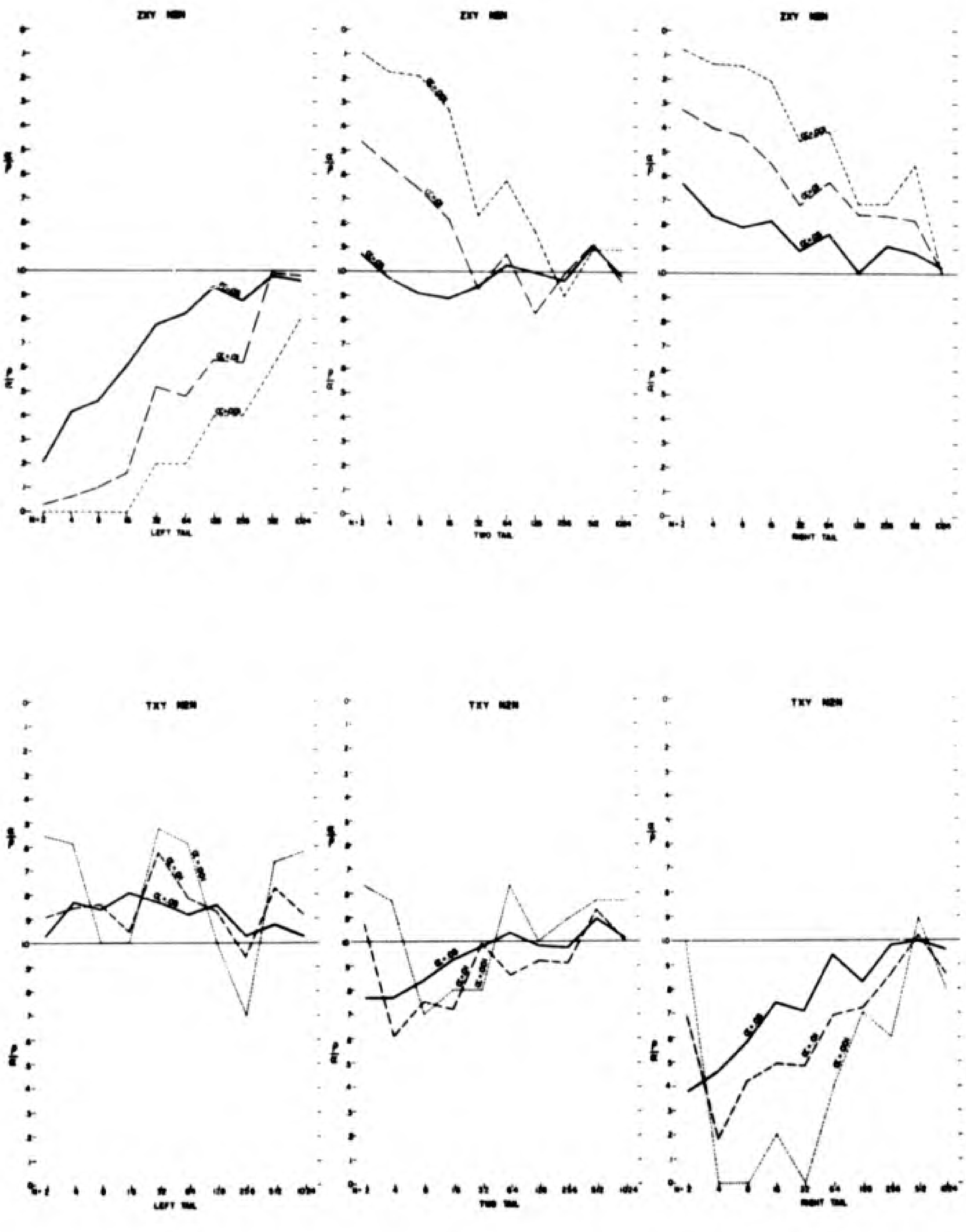

Figure 20. Robustness of ZXY N2N and TXY N2N

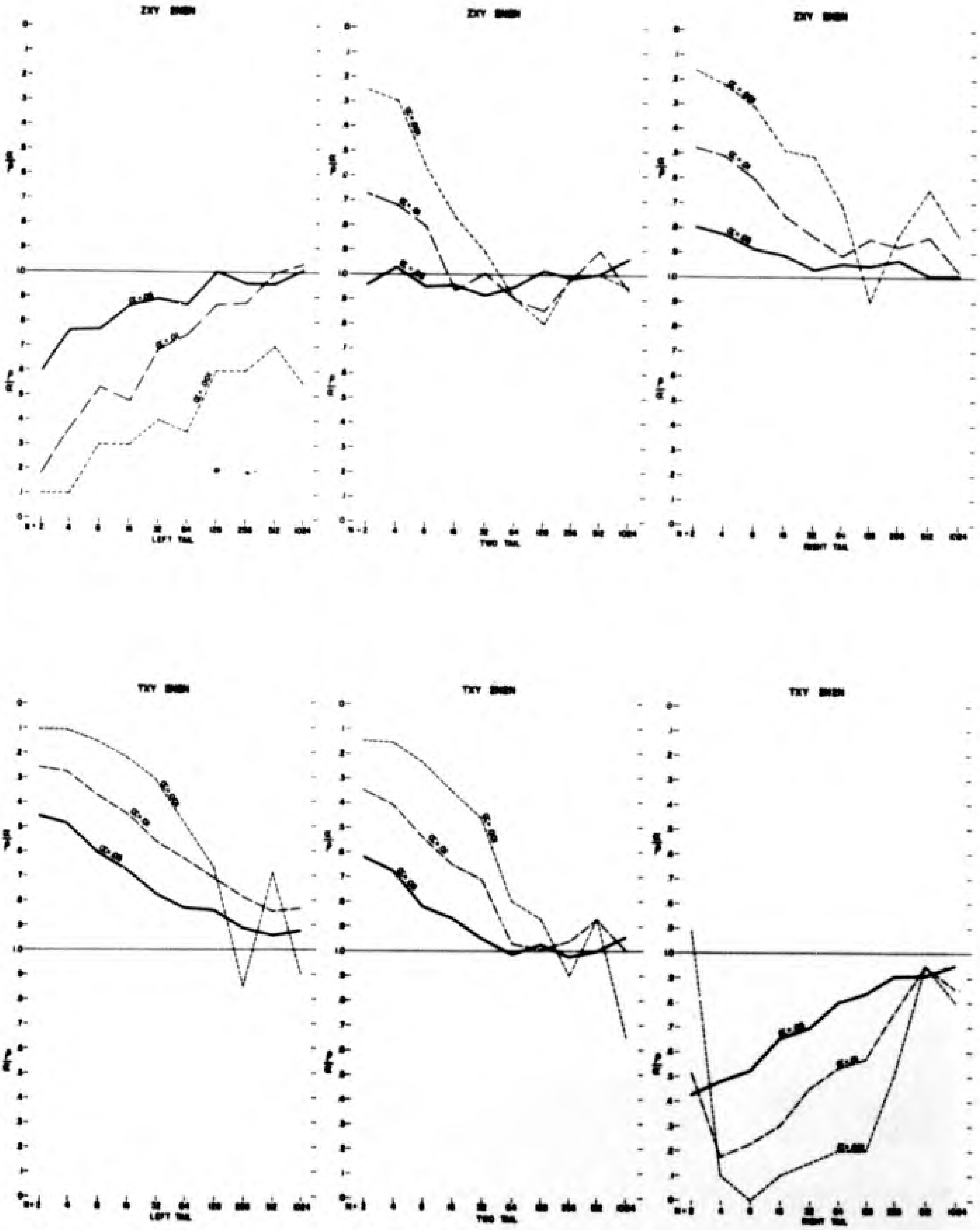

Figure 21. Robustness of ZXY 2N2N and TXY 2N2N

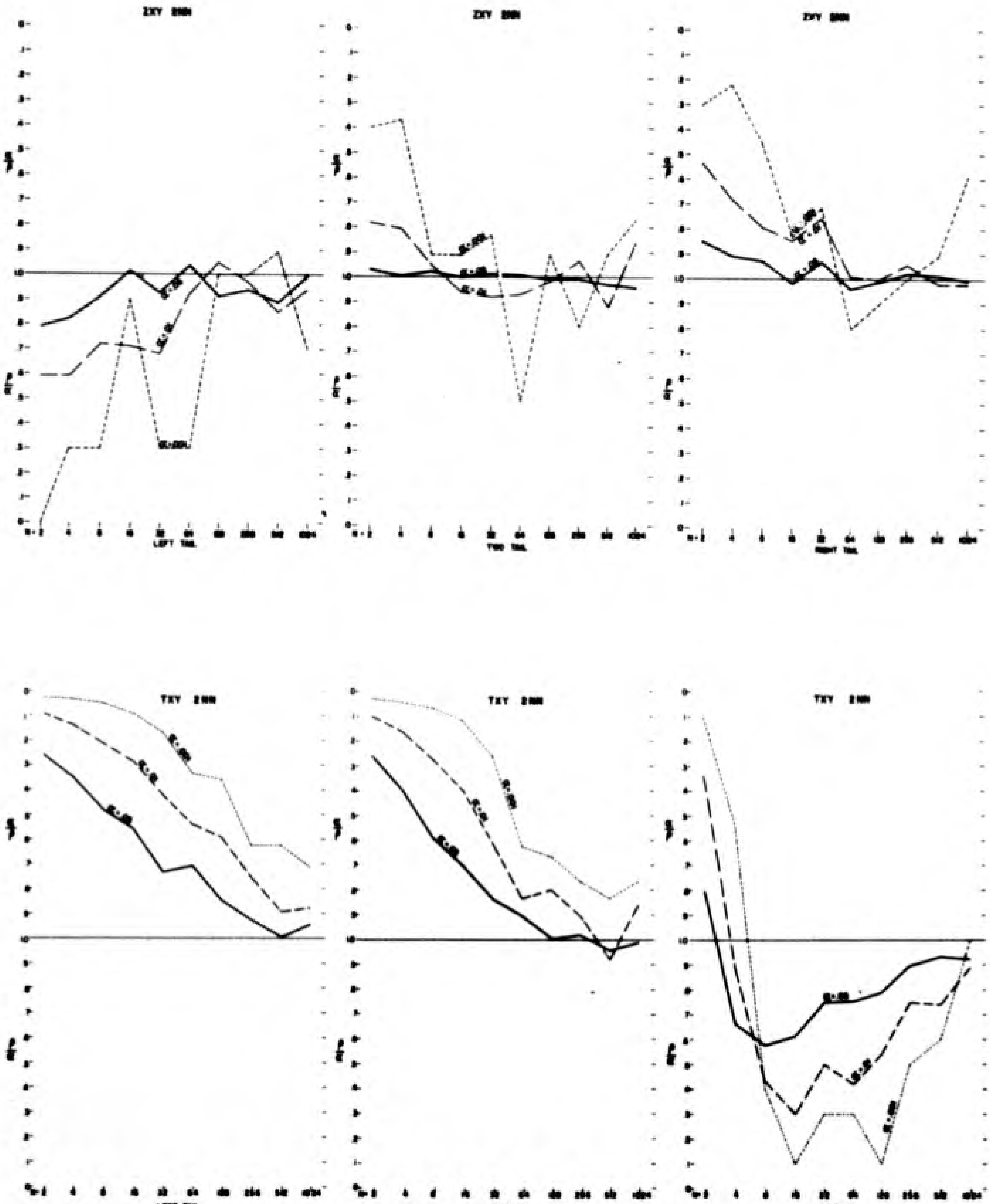

Figure 22. Robustness of ZXY 2NN and TXY 2NN

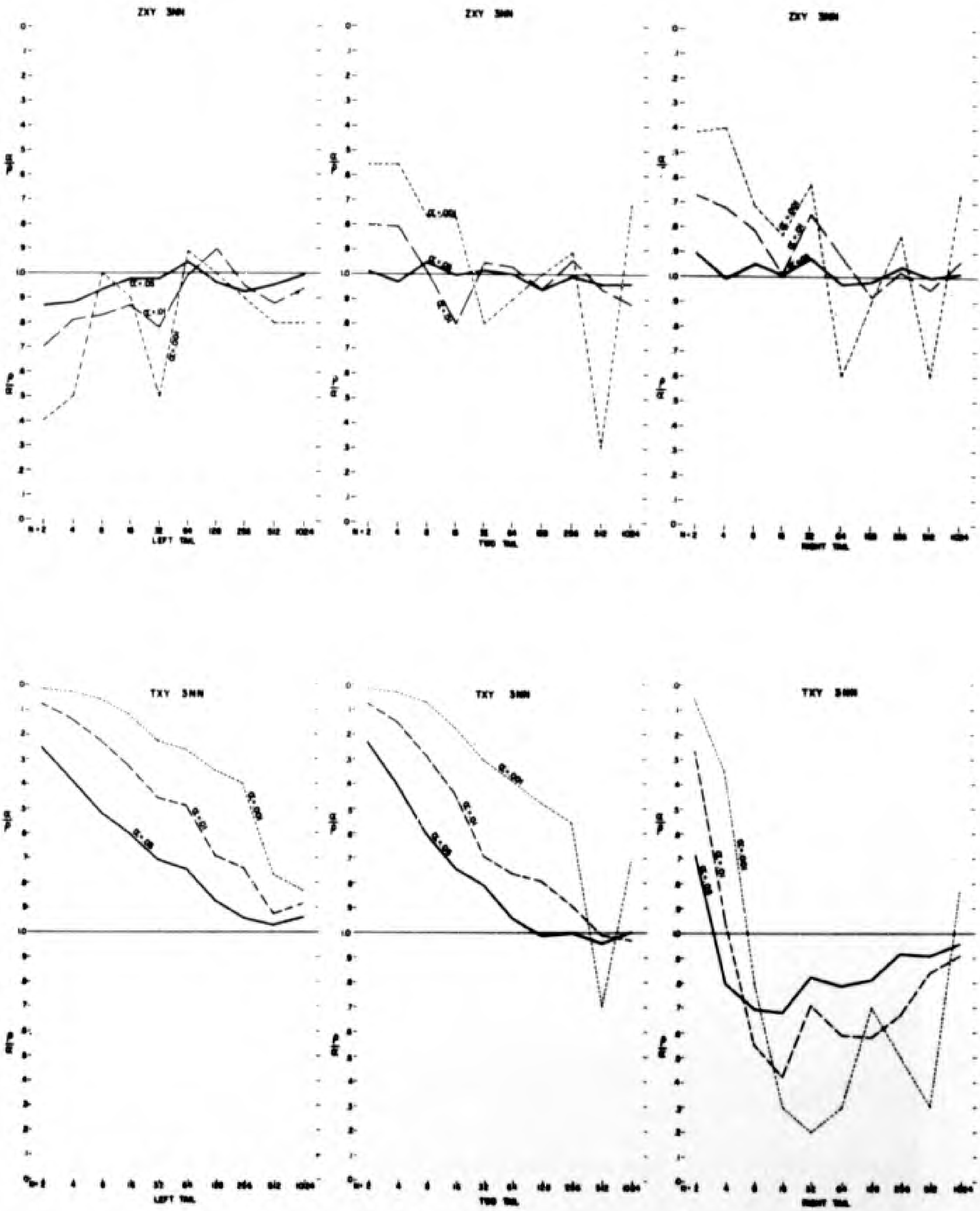

Figure 23. Robustness of ZXY 3NN and TXY 3NN

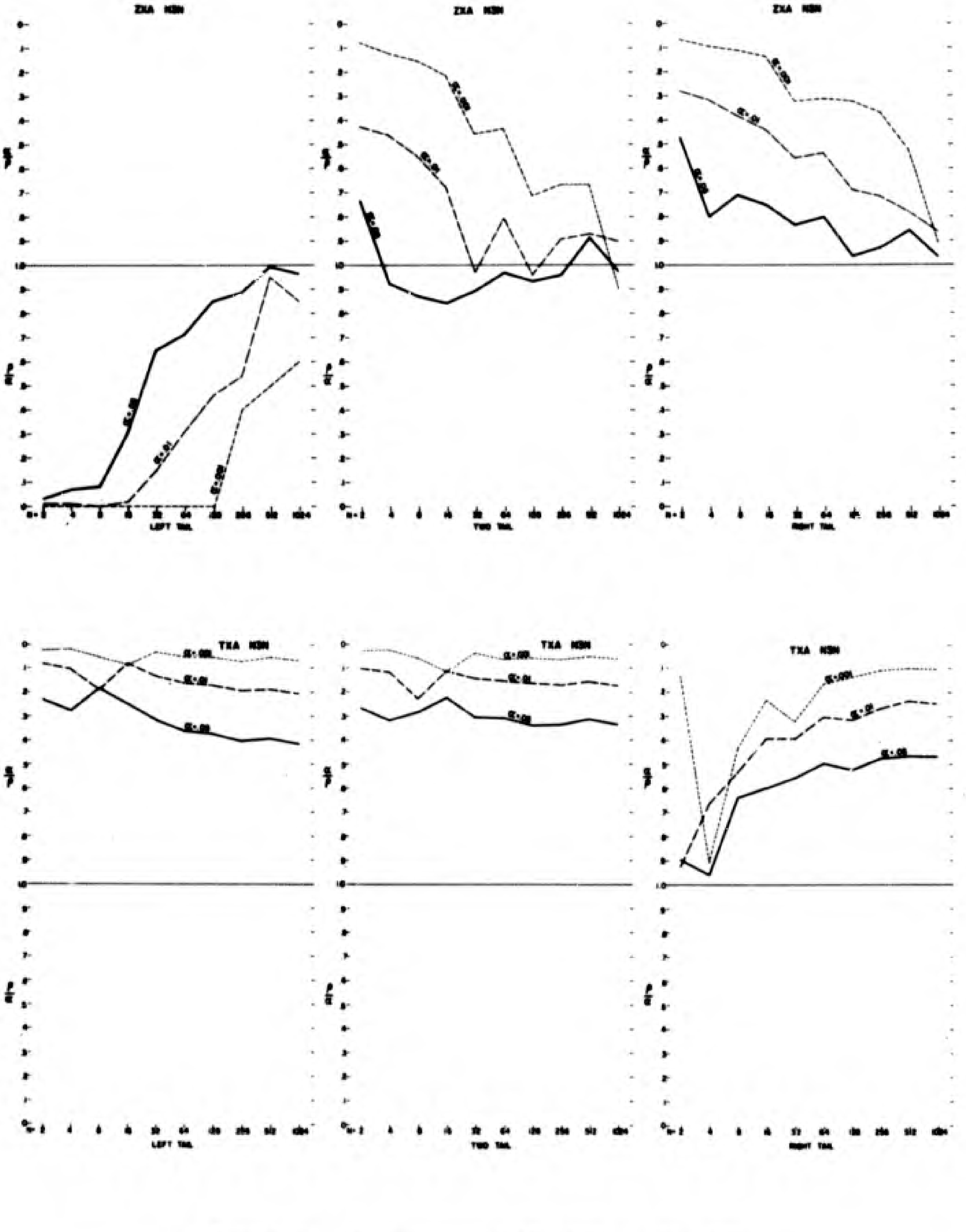

Figure 24. Robustness of ZXA N3N and TXA N3N

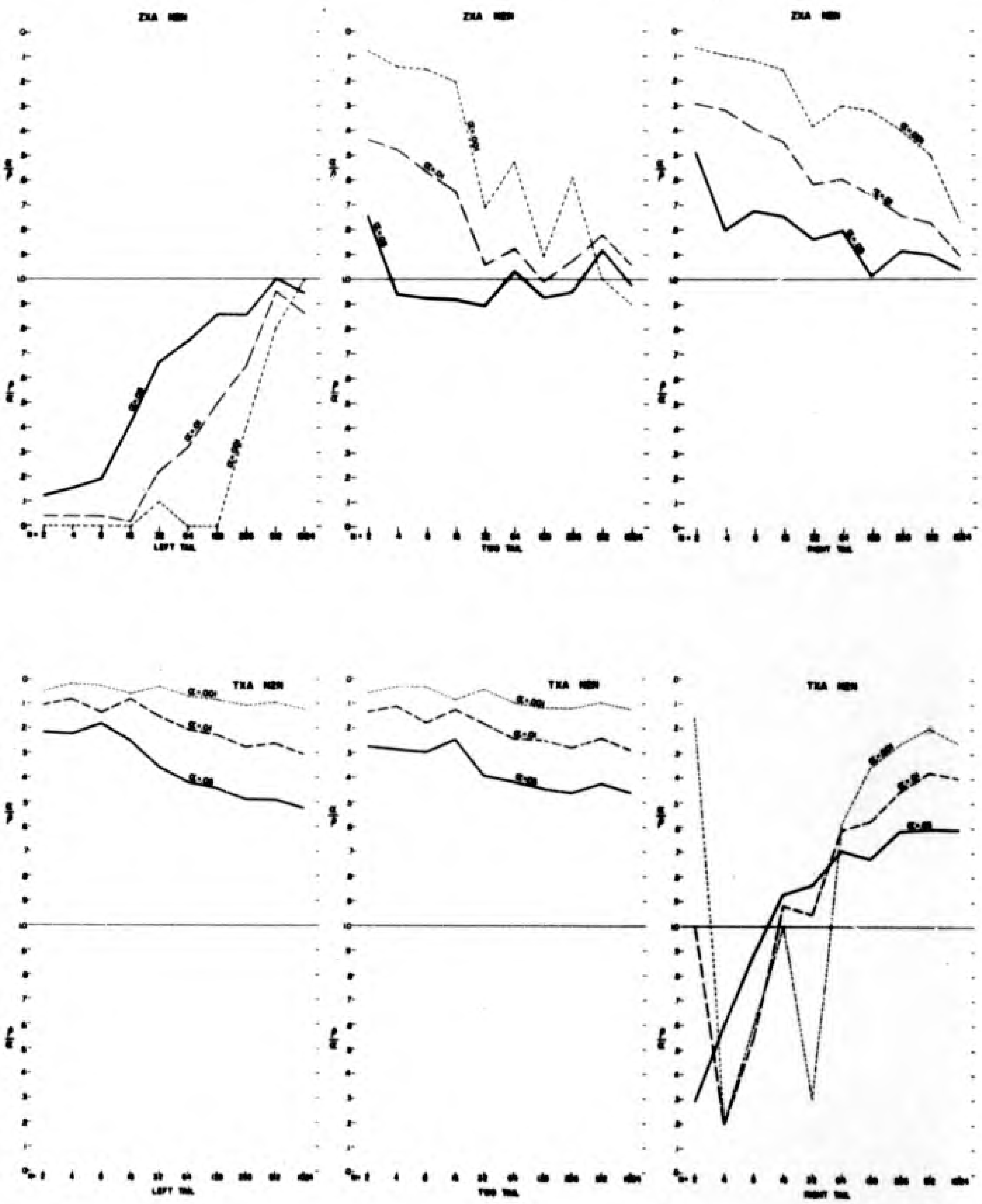

Figure 25. Robustness of ZXA N2N and TXA N2N

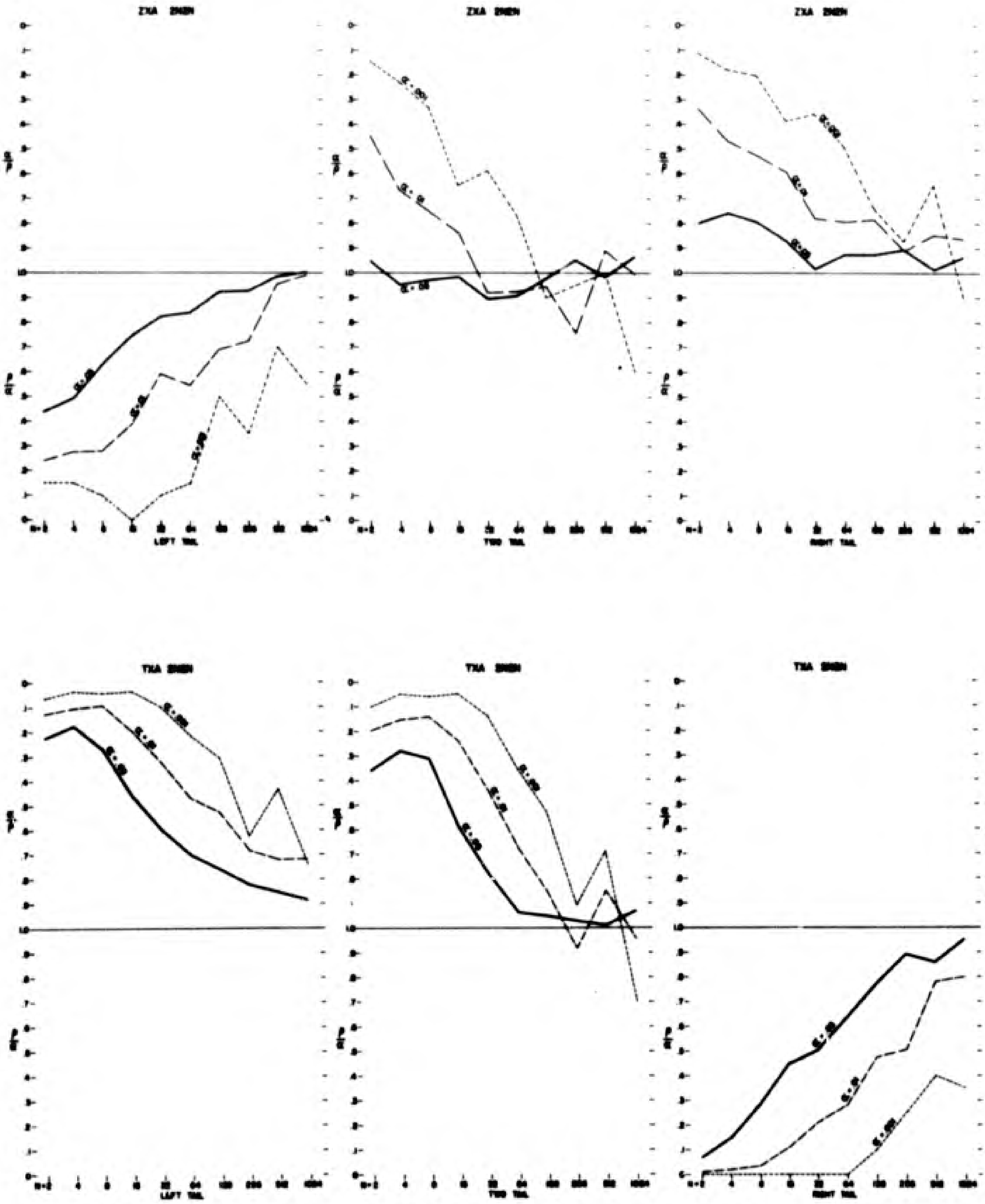

Figure 26. Robustness of ZXA 2N2N and TXA 2N2N

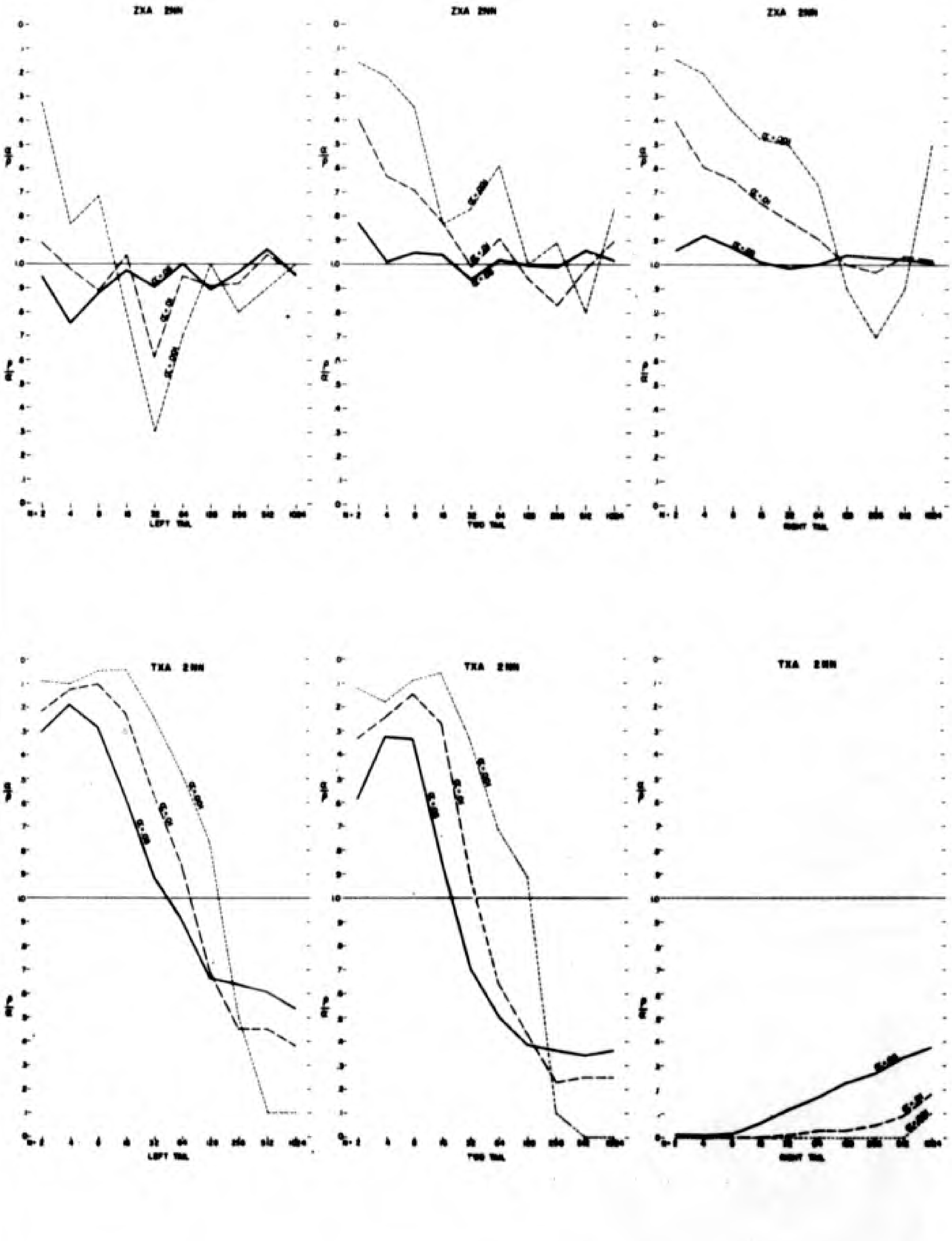

Figure 27. Robustness of ZXA 2NN and TXA 2NN

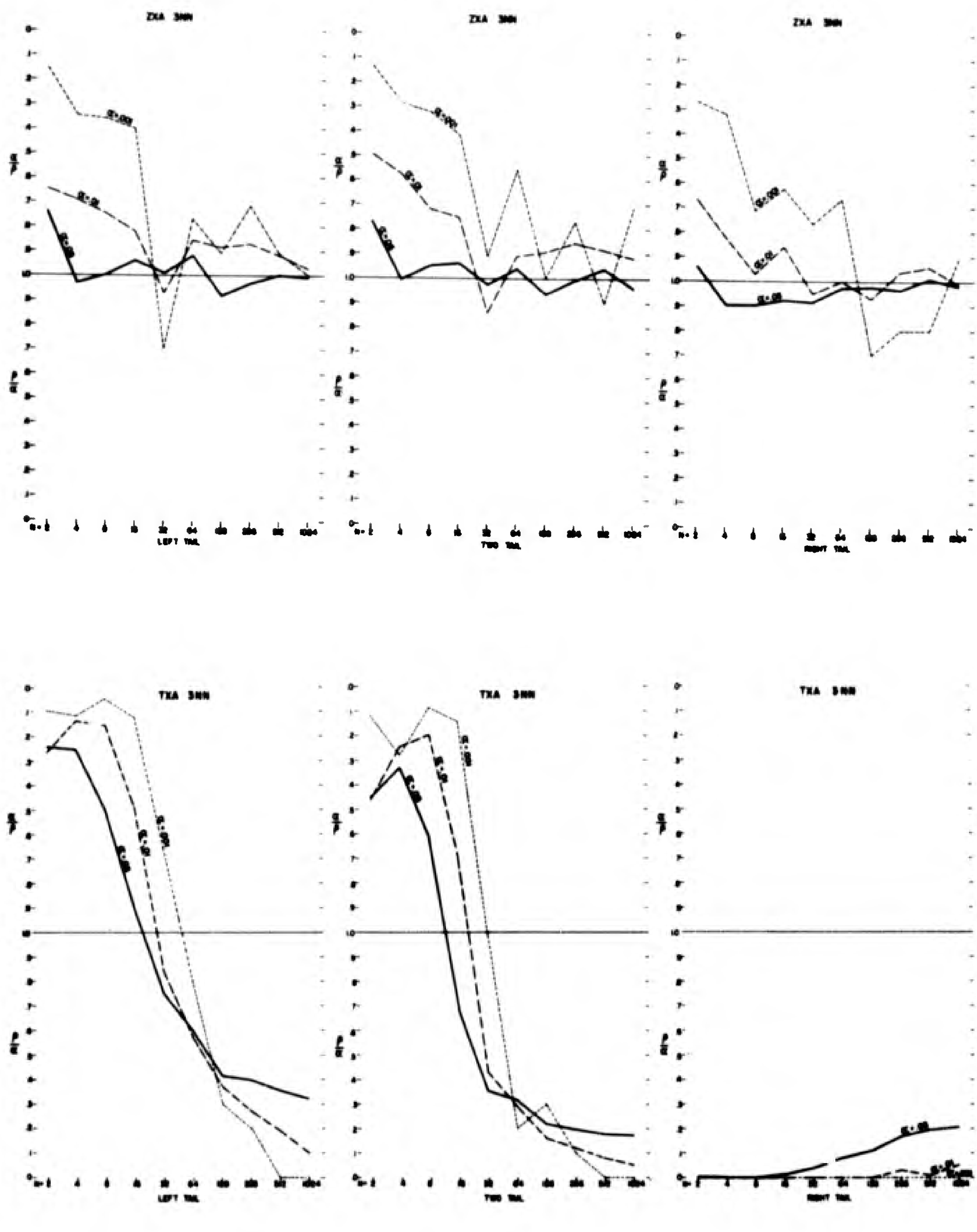

Figure 28. Robustness of ZXA 3NN and TXA 3NN

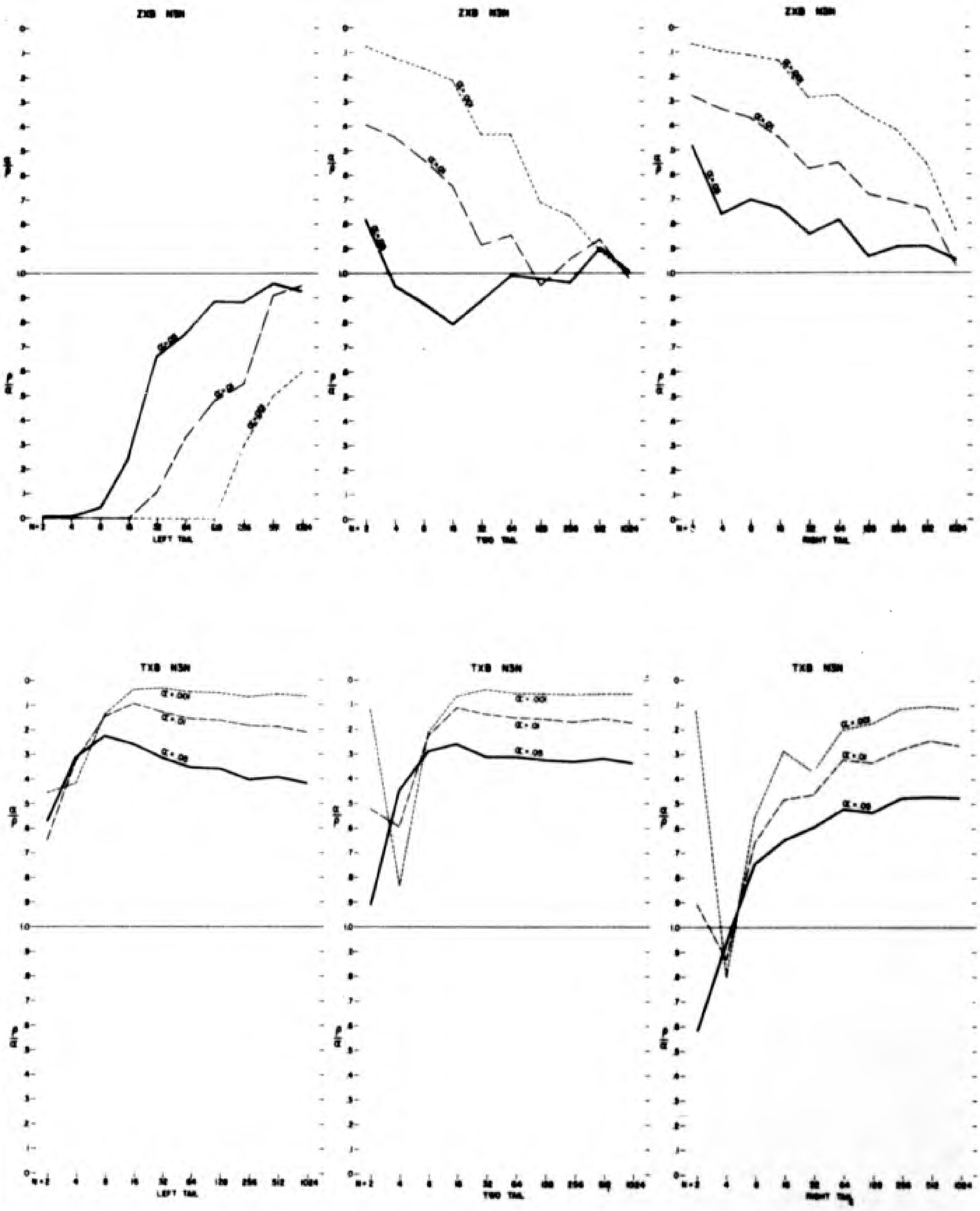

Figure 29. Robustness of ZXB N3N and TXB N3N

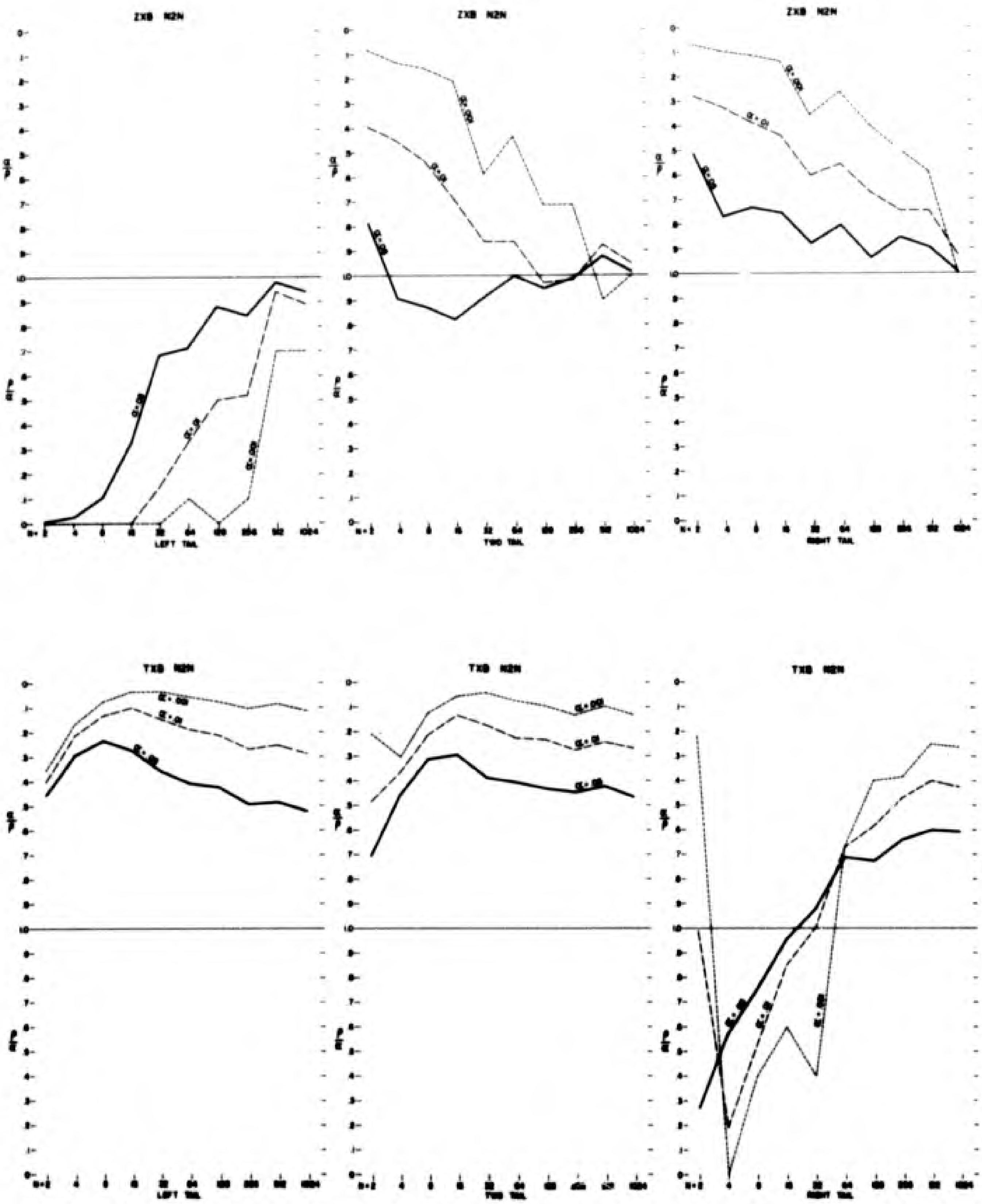

Figure 30. Robustness of ZXB N2N and TXB N2N

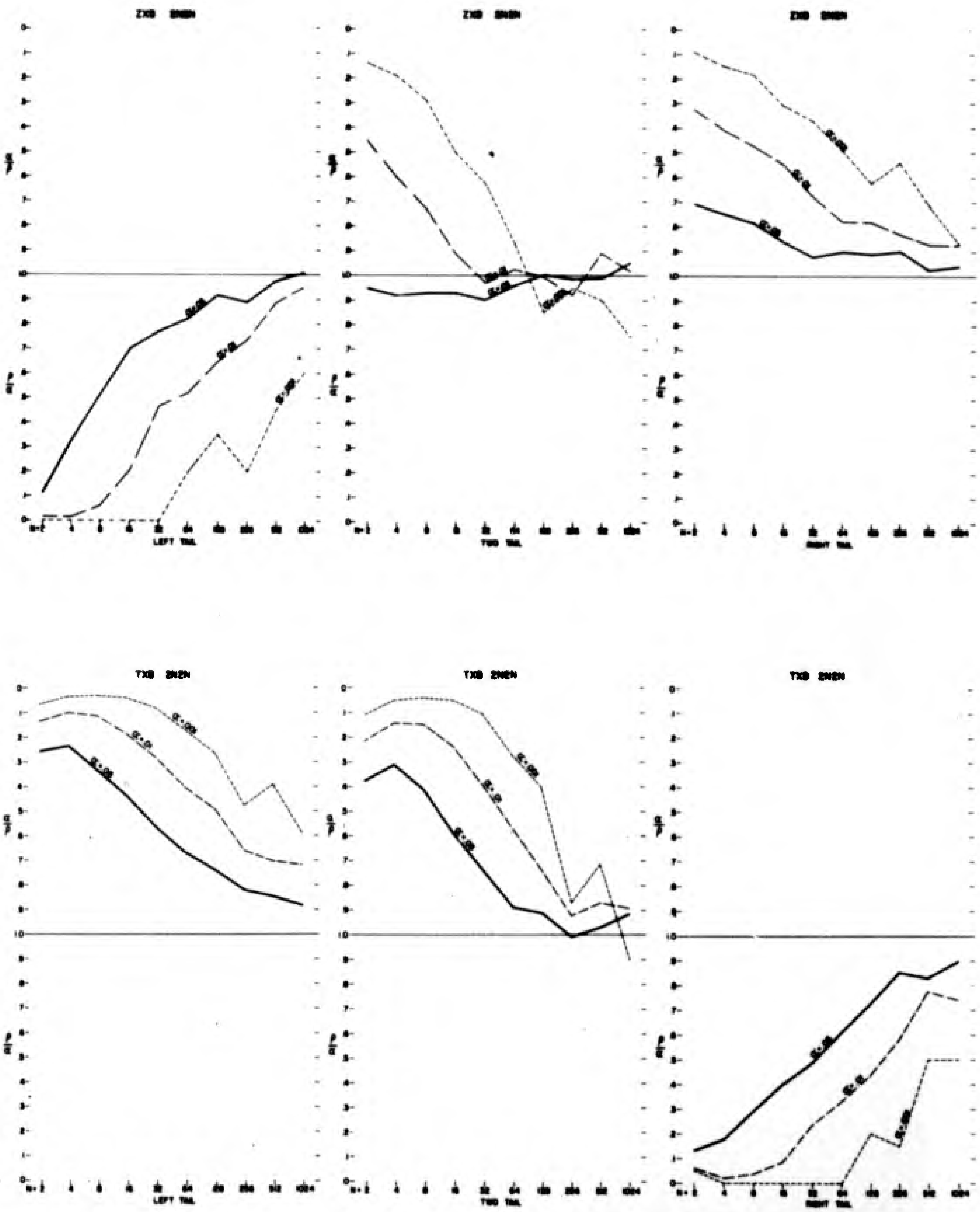

Figure 31. Robustness of ZXB 2N2N and TXB 2N2N

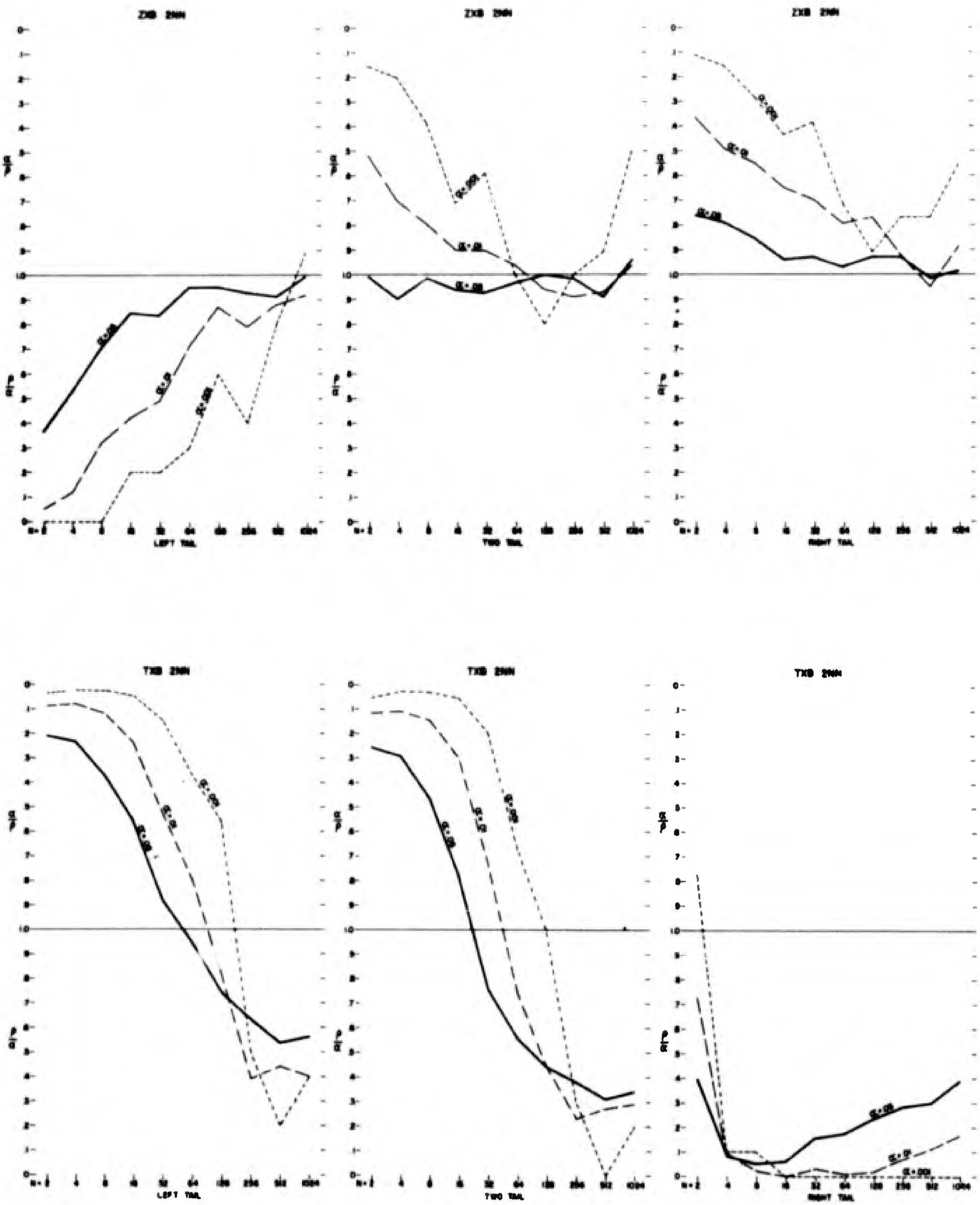

Figure 32. Robustness of ZXB 2NN and TXB 2NN

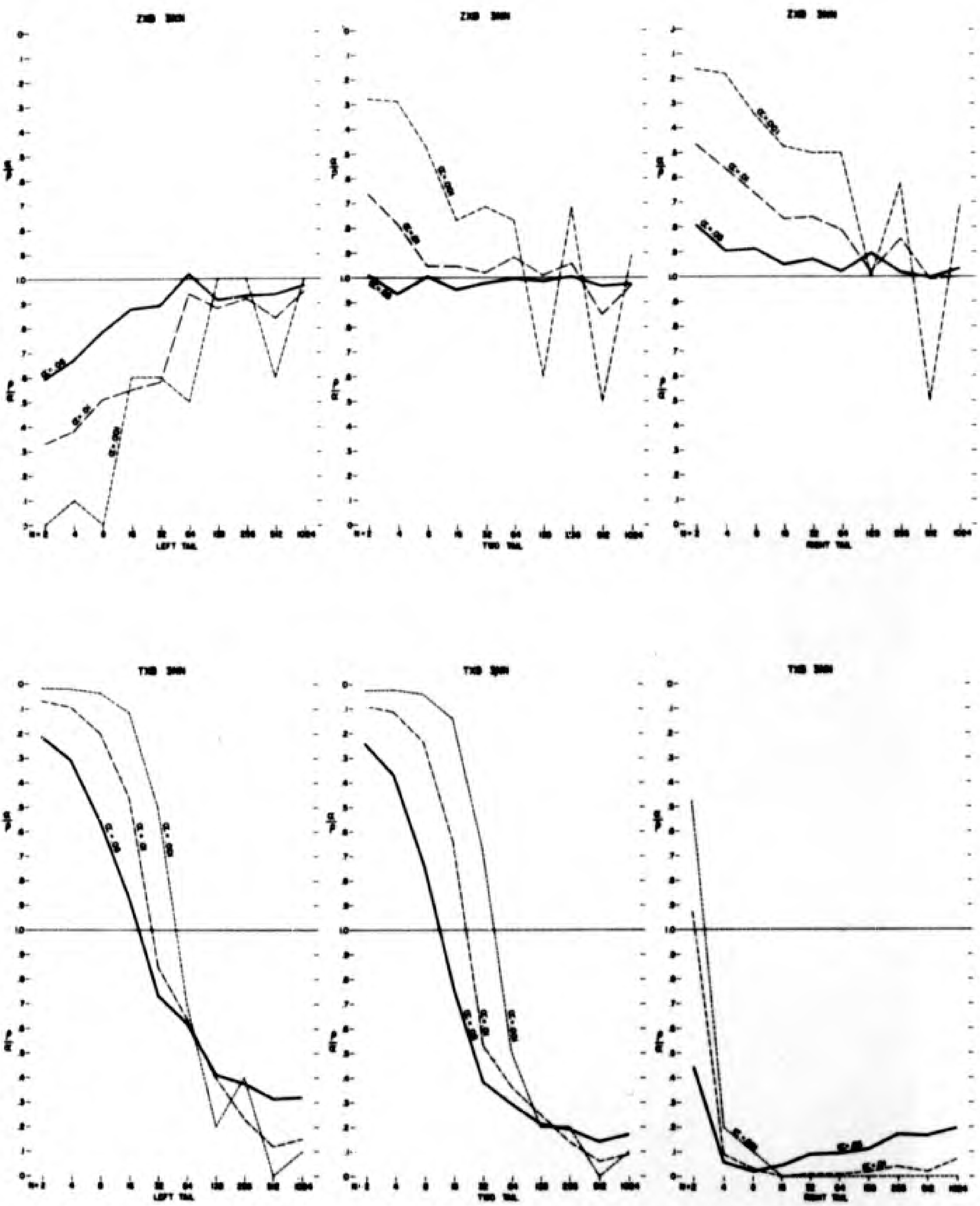

Figure 33. Robustness of ZXB 3NN and TXB 3NN

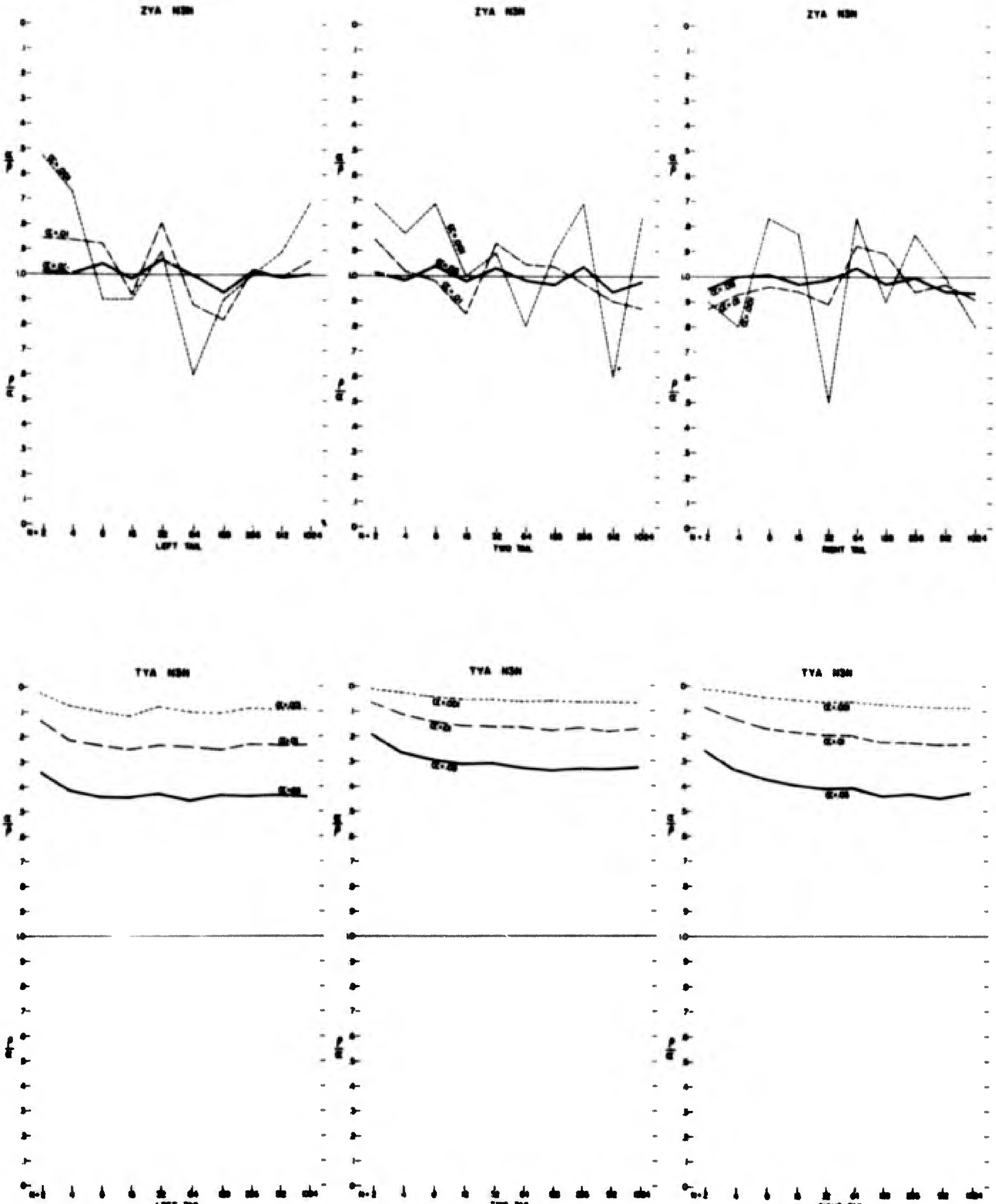

Figure 34. Robustness of ZYA N3N and TYA N3N

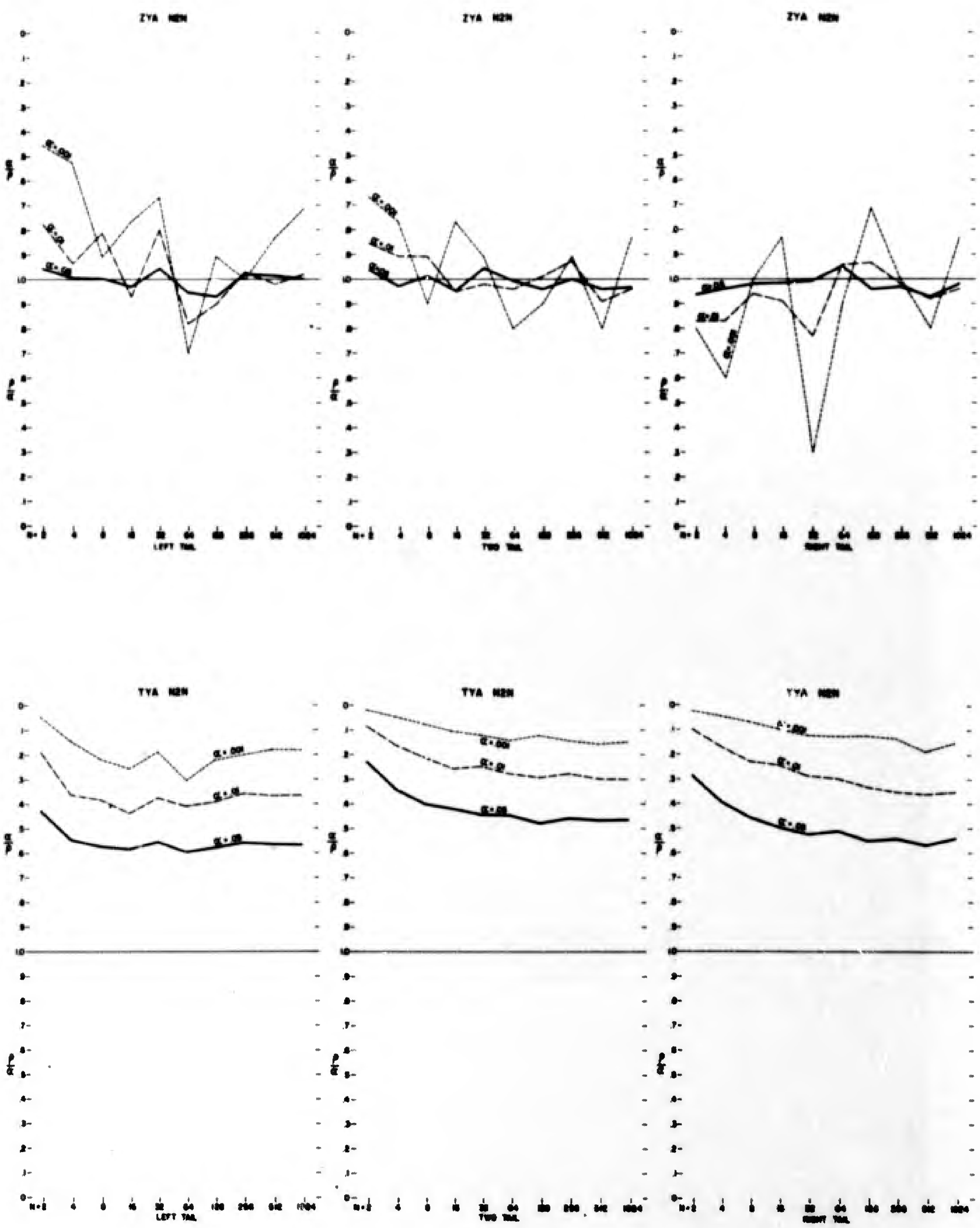

Figure 35. Robustness of ZYA N2N and TYA N2N

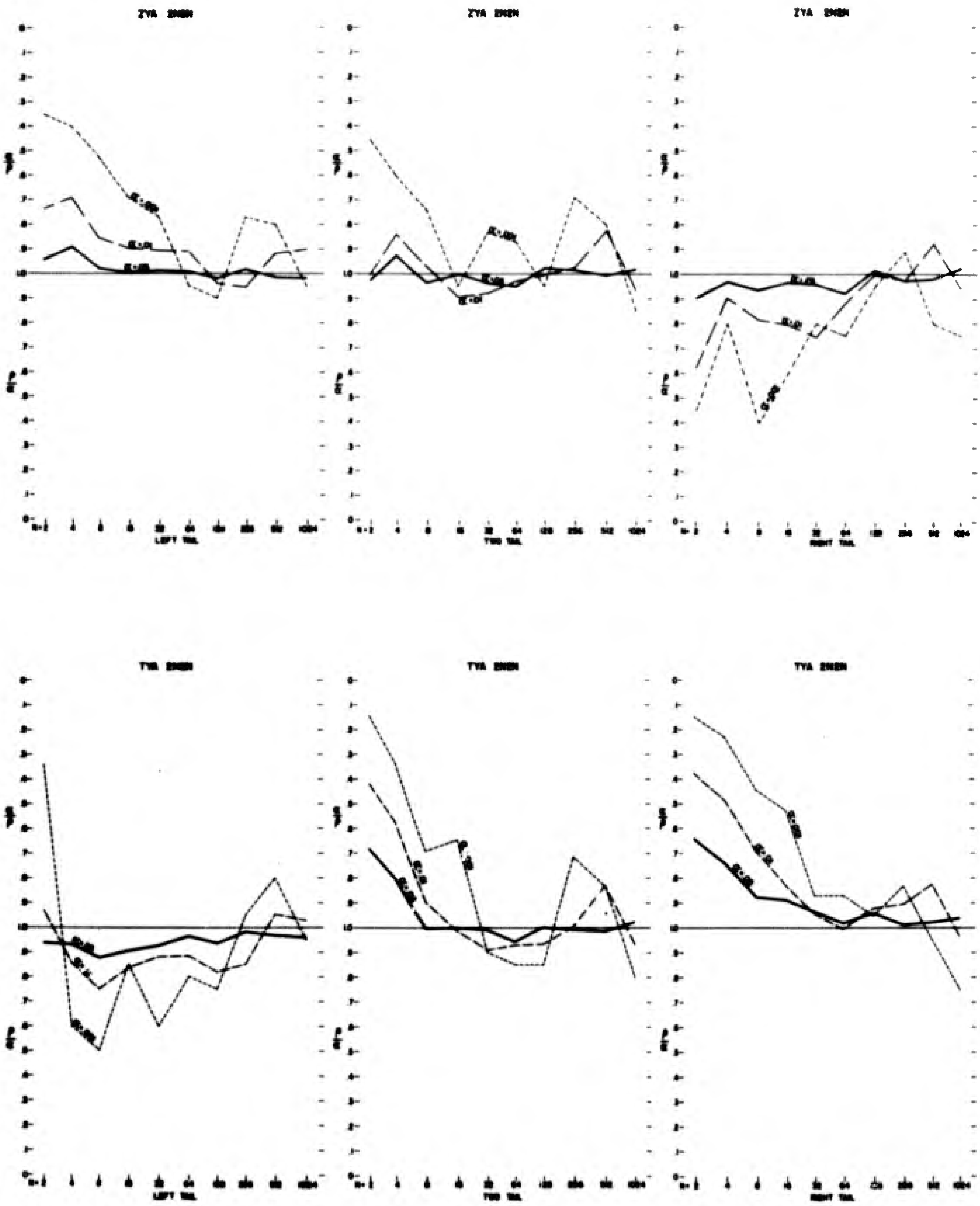

Figure 36. Robustness of ZYA 2N2N and TYA 2N2N

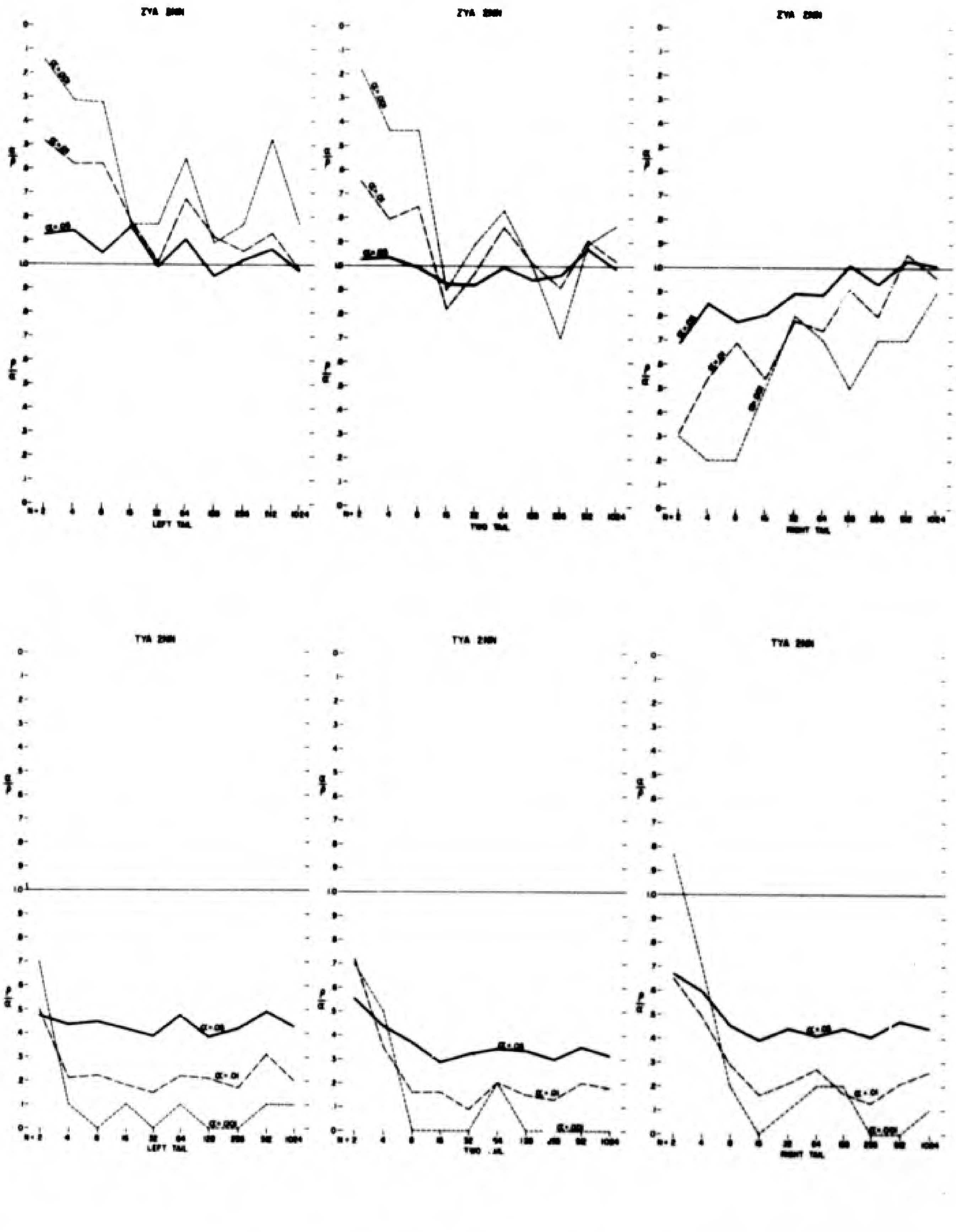

Figure 37. Robustness of ZYA 2NN and TYA 2NN

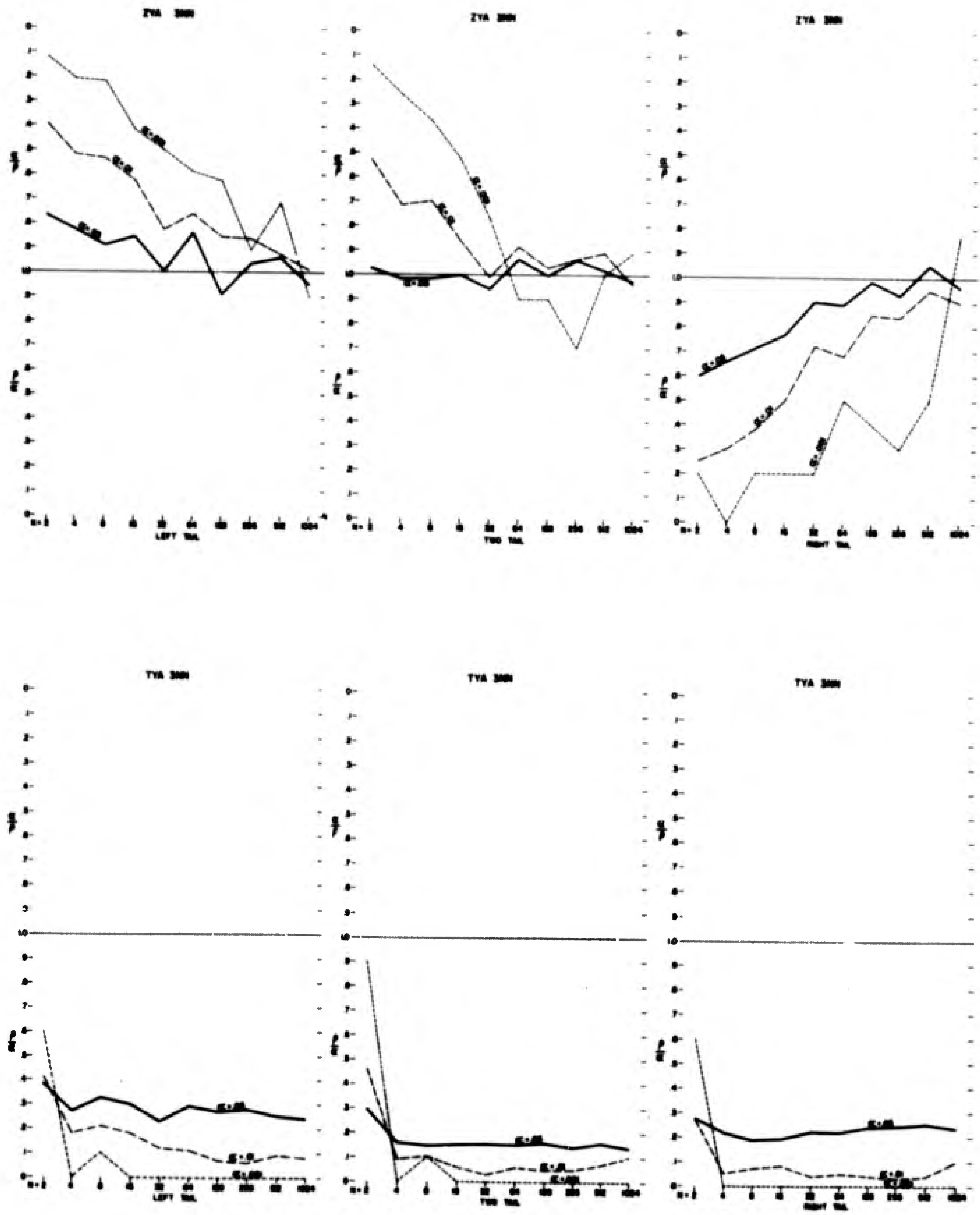

Figure 38. Robustness of ZYA 3NN and TYA 3NN

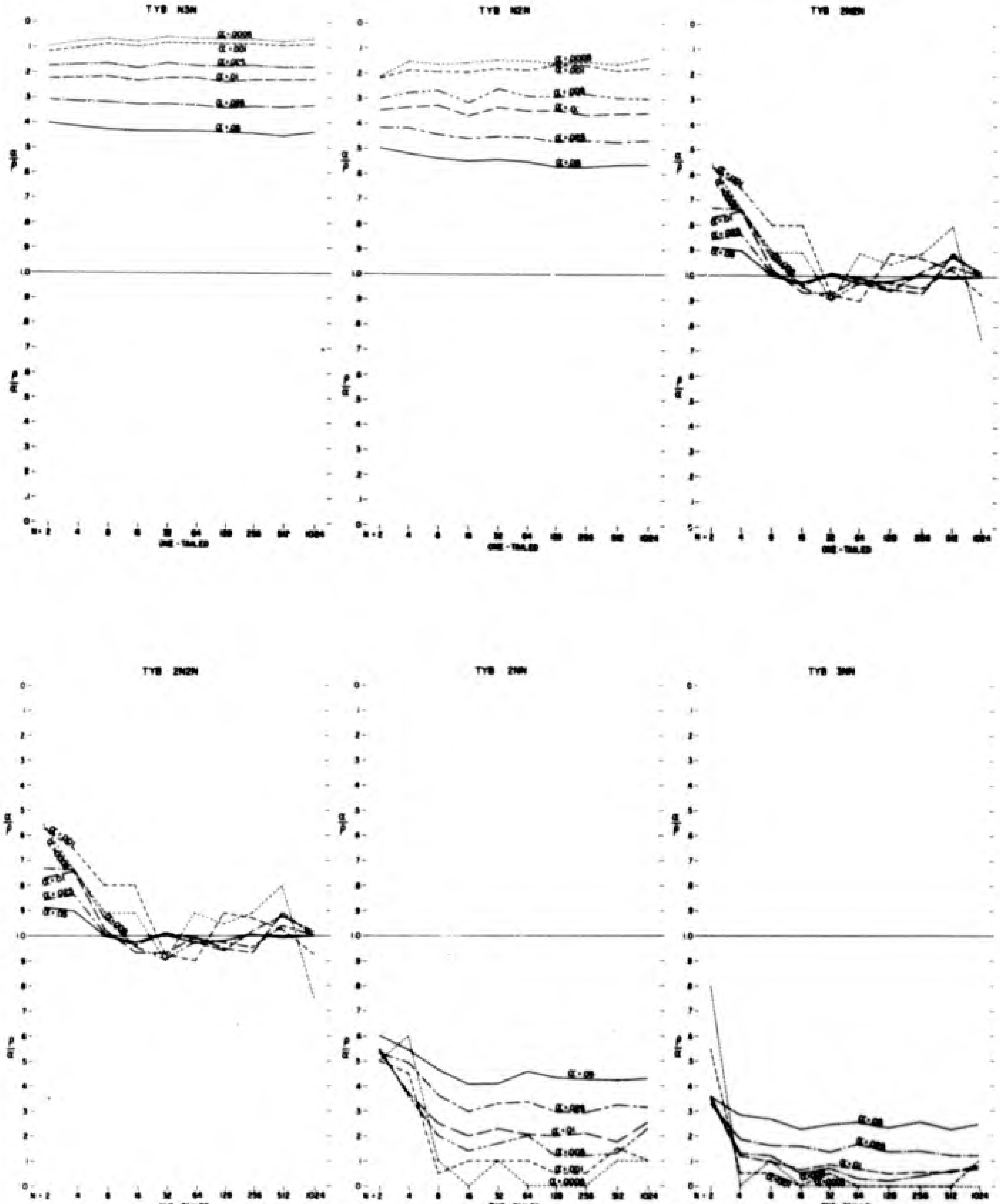

Figure 39. Robustness of TYB

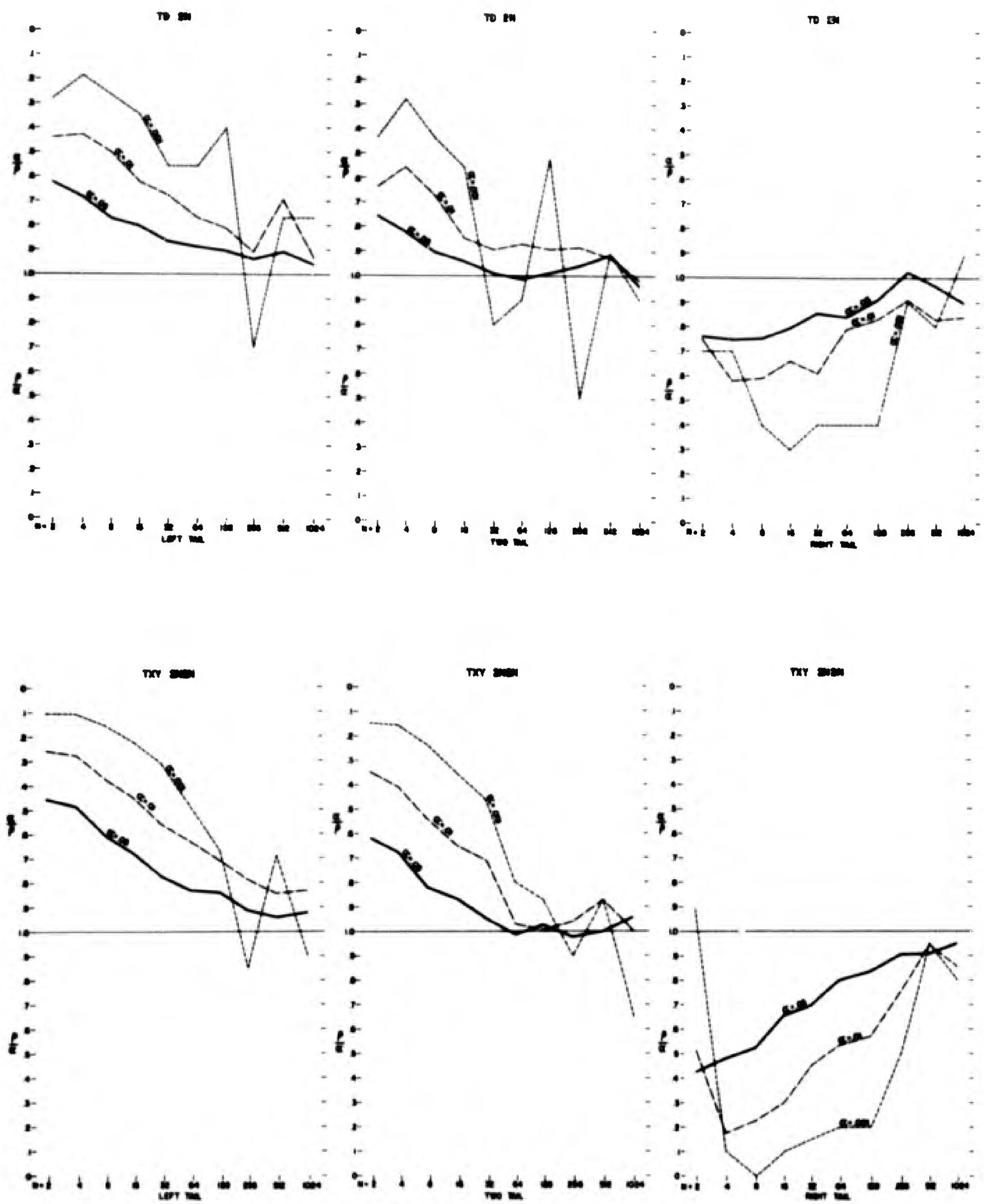

Figure 40. Robustness of TD 2N and (repeated graphs) TXY 2N2N

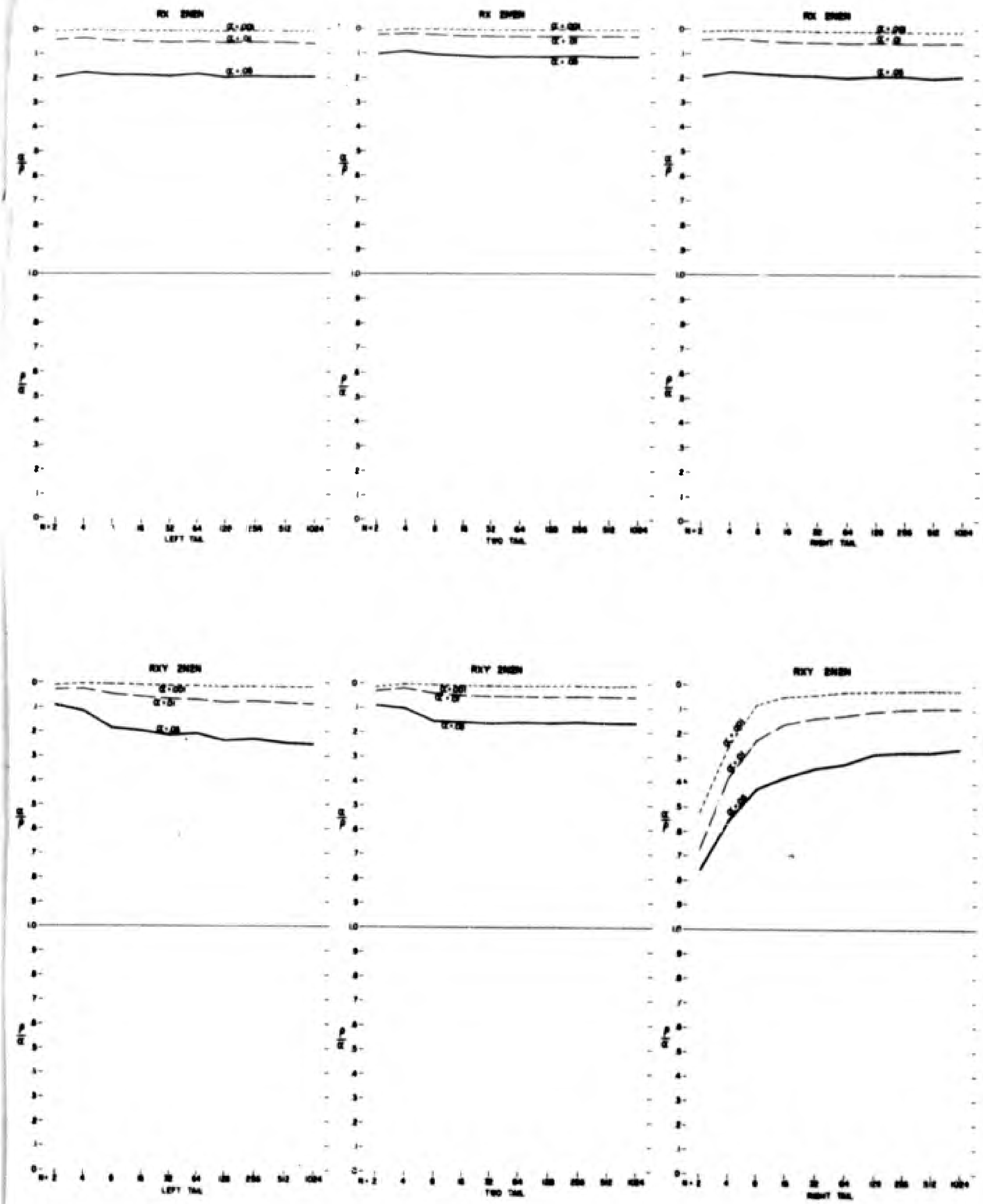

Figure 41. Robustness of RXX 2N2N and RXY 2N2N

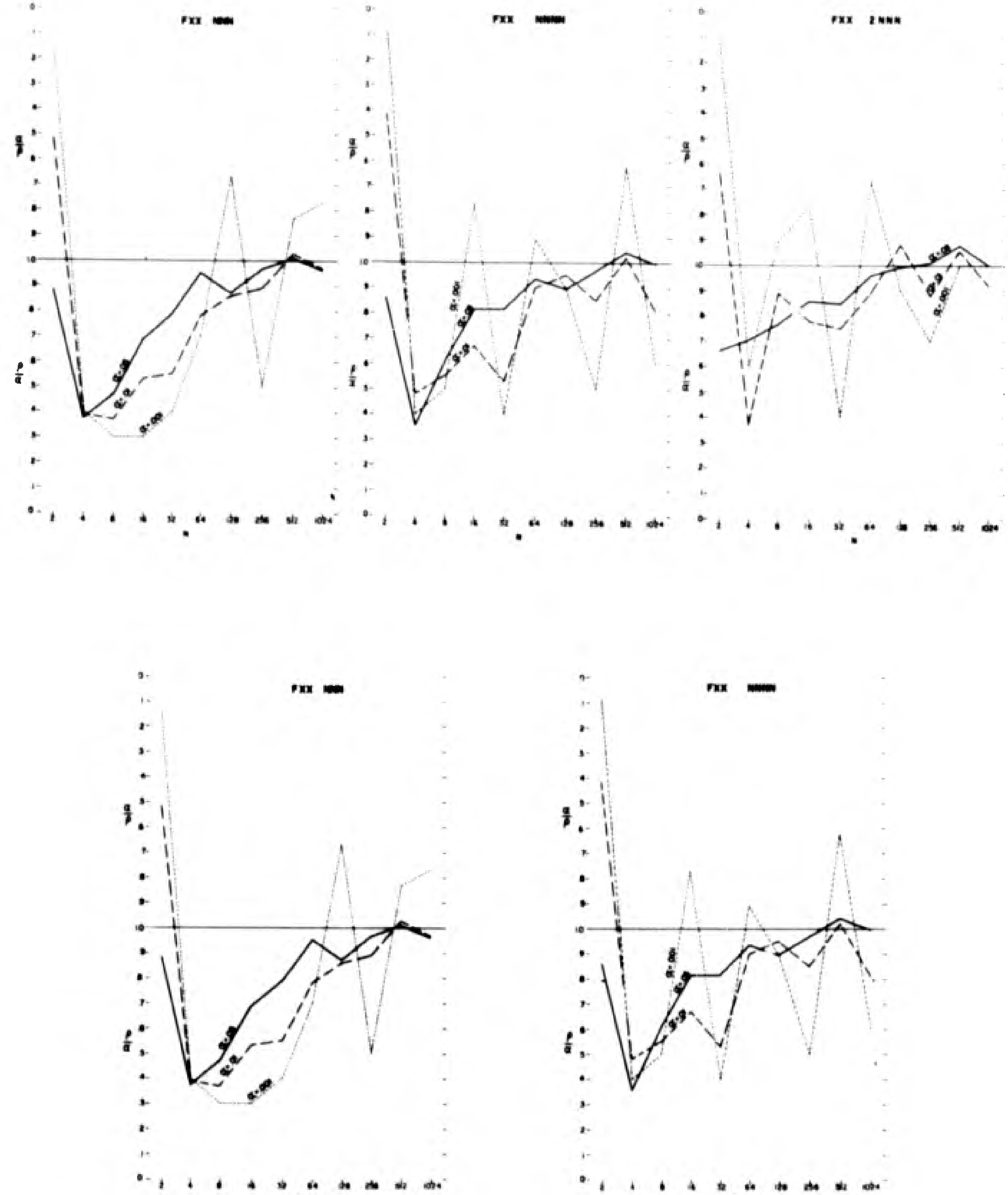

Figure 42. Robustness of FXX (lower graphs are repeats)

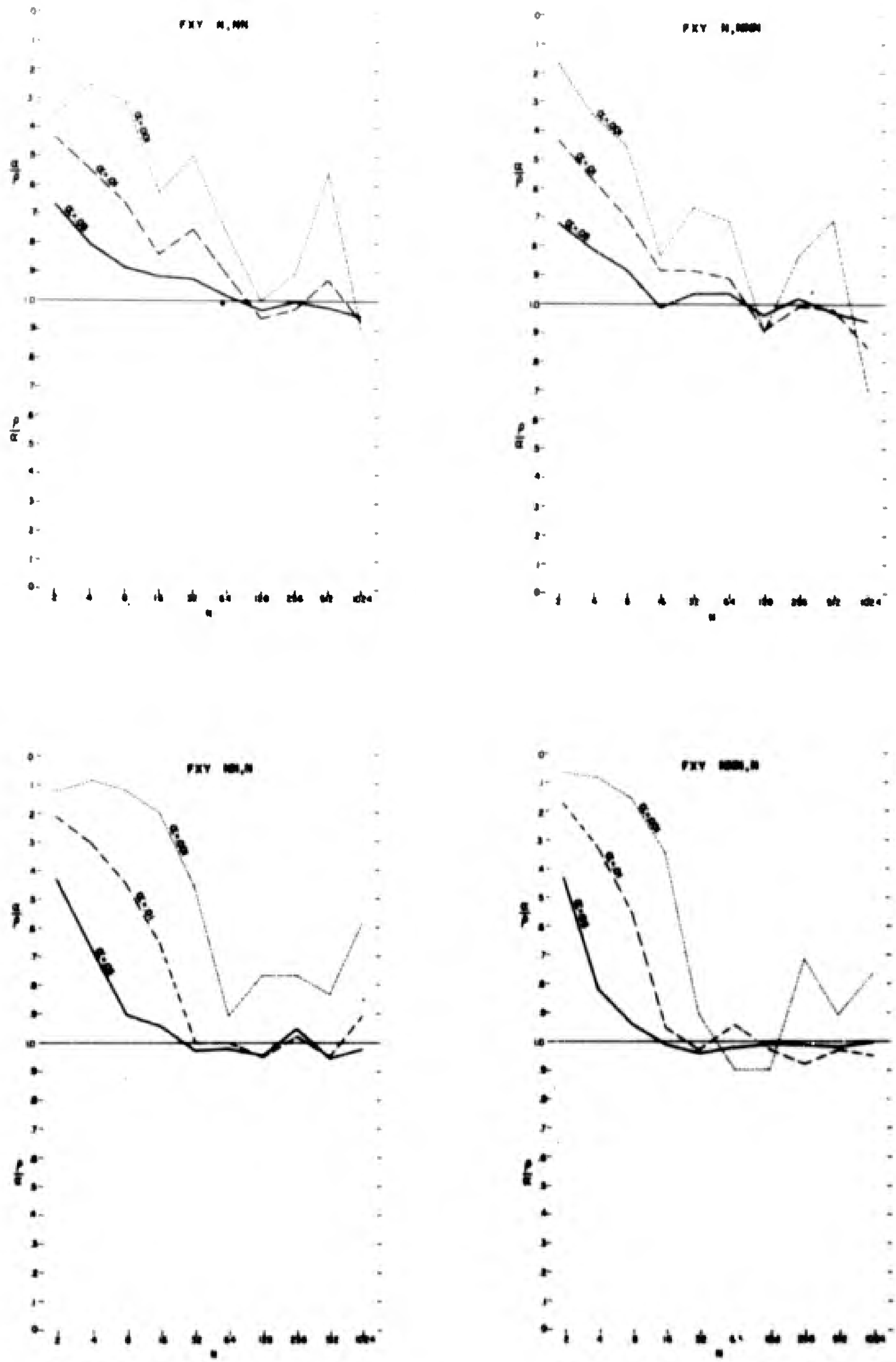

Figure 43. Robustness of FXY

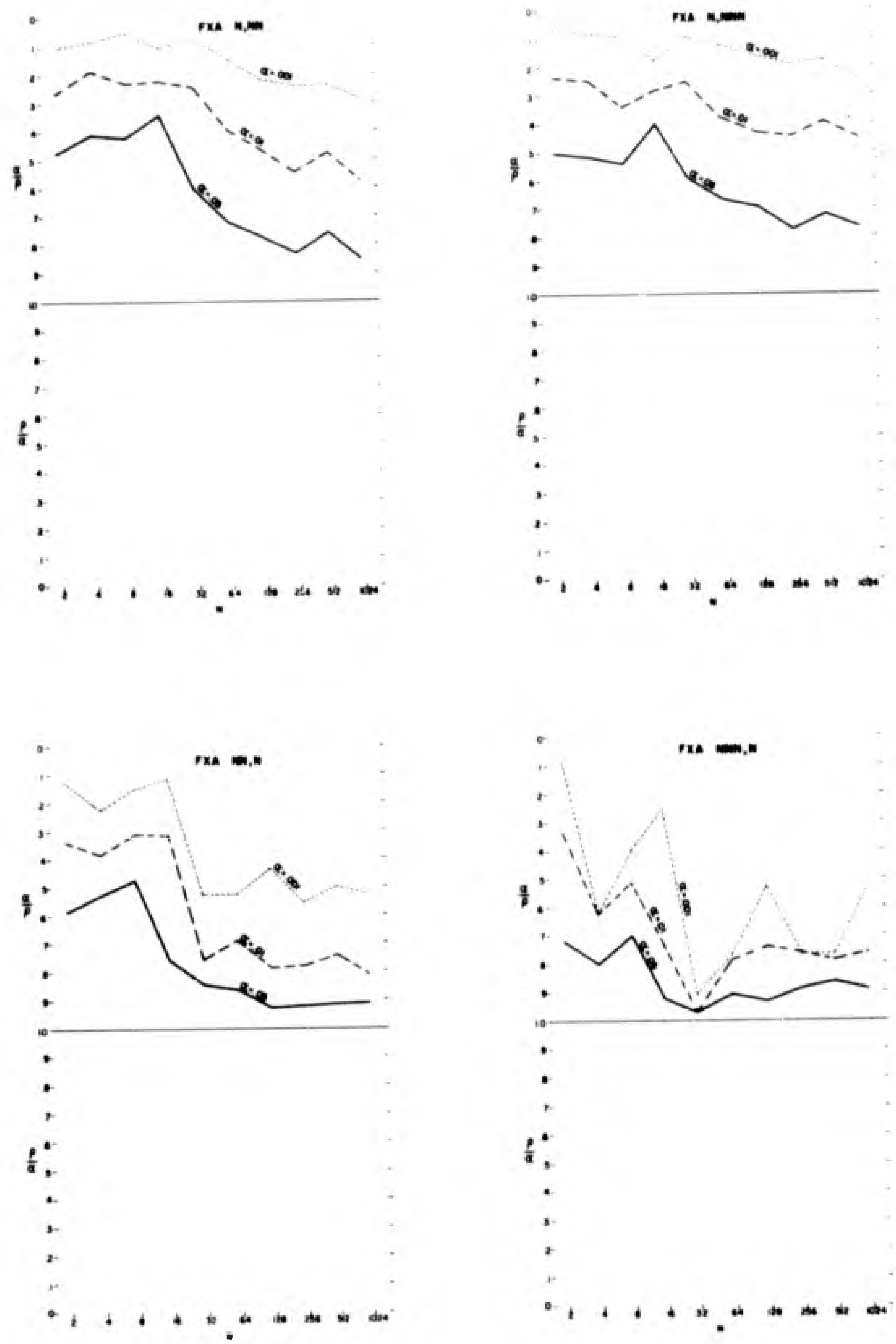

Figure 44. Robustness of FXA

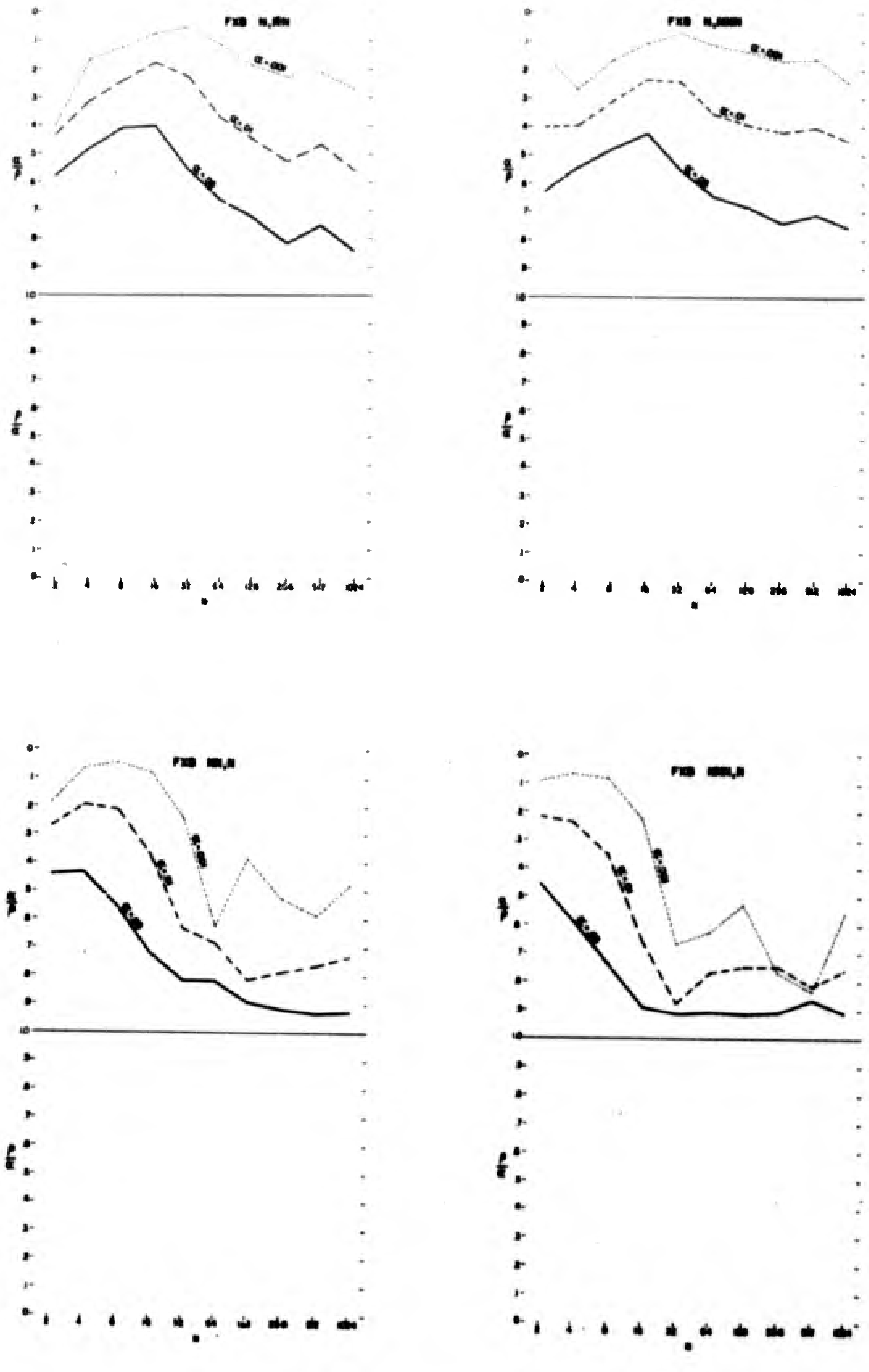

Figure 45. Robustness of FXB

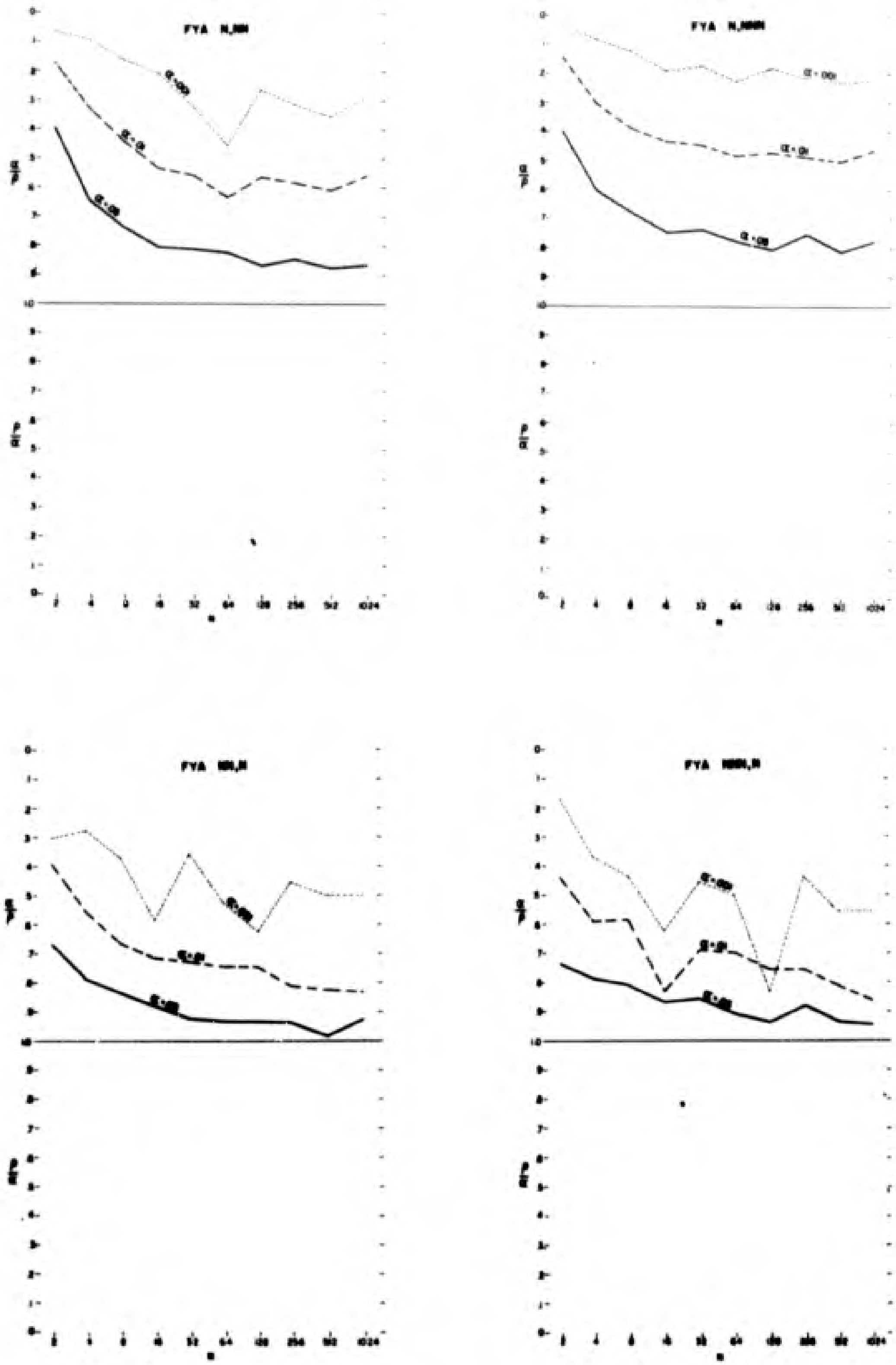

Figure 46. Robustness of FYA

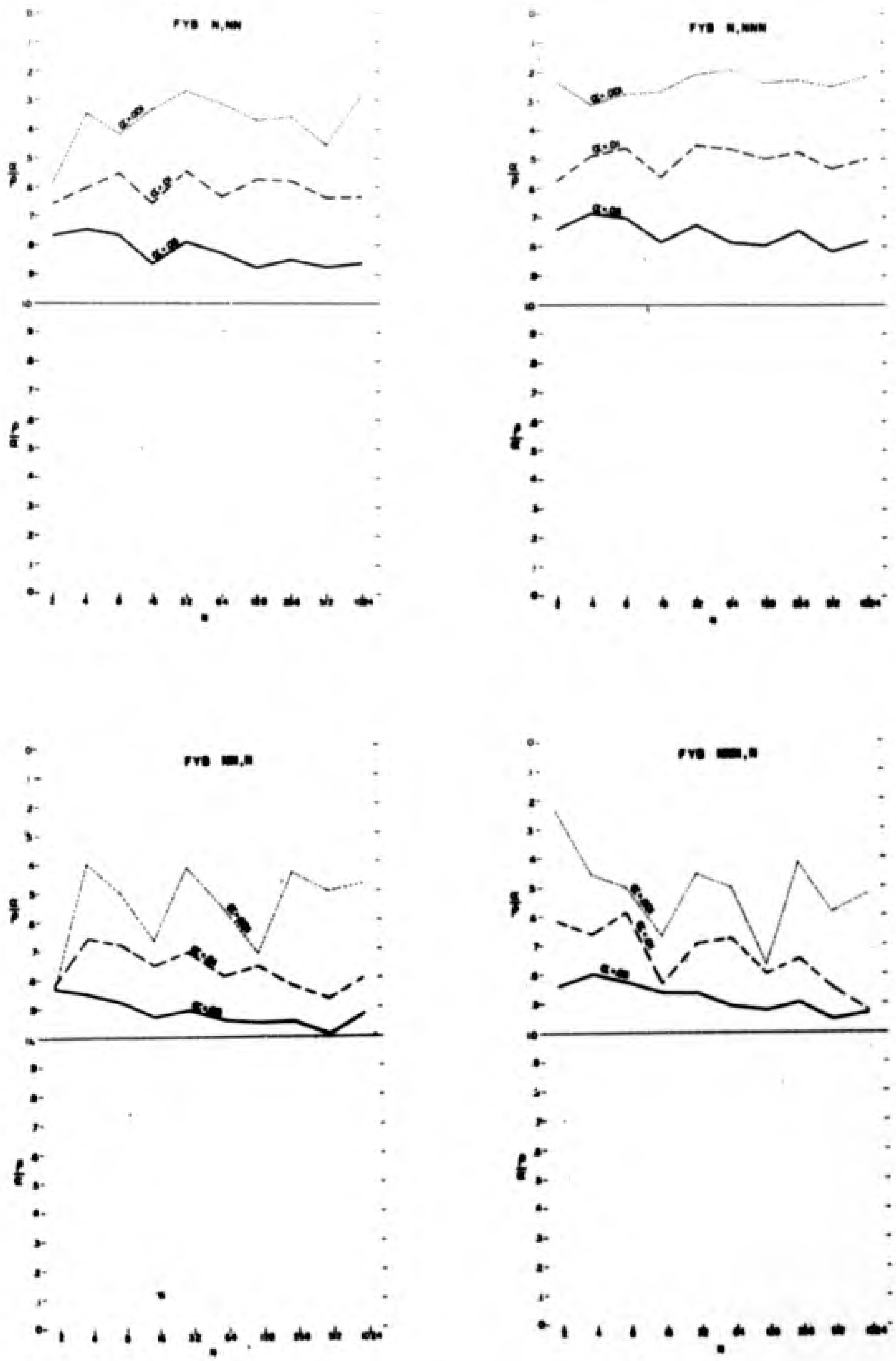

Figure 47. Robustness of FYB

DISCUSSION

I The Central Limit Effect

The efficacy of the Central Limit Effect in producing quasi-normality in the distribution of the sample mean can be measured in terms of the relative discrepancy between ρ and α, i.e., in terms of robustness. In order to have an objective measure of this efficacy, an index of robustness was devised. For each of the six relevant Z statistics, $\frac{|\rho-\alpha|}{\alpha}$ was obtained at a sample size of 16 (i.e., at N = 16 for ZX and at 2N = 16 for ZXX 2N2N, ZXY 2N2N, ZXA 2N2N, ZXB 2N2N and ZYA 2N2N) for both left-tailed and right tailed α's of both .05 and .01. The average of these four values of $\frac{|\rho-\alpha|}{\alpha}$ was defined to be the index of nonrobustness (of the sample mean) for the population involved, and its complement was defined to be the index of robustness. When arranged in order of increasing robustness on the basis of these indices, the populations appear in the order X, X-B, X-A, X-Y, Y-A, X-X (and their respective indices of robustness are .072, .295, .437, .629, .890, and .935). One might suppose that the more closely the sampled population is fitted by a normal distribution with the same mean variance and area, the closer will be the fit between the distribution of the sample mean and the normal distribution which the Central Limit Theorem says it approaches. In order to have an objective measure with which to explore this possibility an index of normality was defined to be the proportion of the total area of a nonnormal population which is common to a normal distribution with the same mean, variance, and area. The required proportions were obtained, i.e., measured, by taking the photographed figures showing the nonnormal population histograms and accompanying normal curves, cutting out of the photograph the area corresponding to the normal curve (which equals the area of the histogram), weighing it in a chemical balance, then cutting out of the normal area that portion of it which was common to the histogram and weighing that portion. The ratio of the latter weight to the former is the required index of normality, and its complement is the index of nonnormality. In order of increasing indices of normality the populations are X, X-A, X-X, X-B, X-Y, Y-A (and their respective indices of normality are .305, .406, .408, .668, .861, and .969). In order to convert this sequence into that for increasing robustness, X-X would have to be moved to the extreme right and the positions of X-A and X-B would have to be switched. If this were done, the populations would be in order of apparently increasing symmetry, i.e., increasing subjectively-estimated symmetry. Further evidence suggests that robustness is far more strongly influenced by the symmetry-vs.-asymmetry dimension than by the closeness-to-normality factor. The X-X and X-A populations are quite similar in the proportional amount of their area which lies under the normal curve and also in general shape. However, means from the exactly symmetrical X-X population are far more robust than are those from X-A. The Y-A population is very closely fitted by the normal curve and except for a highly improbable tail portion is very nearly symmetrical. However, its sample means are less robust than those from the X-X population which has one of the smallest proportions of area in common with the normal curve but which is exactly symmetrical.

The evidence, therefore, favors the conclusion that symmetry is a more important factor than quasi-normality in producing robustness. The inability of the trimodality, long tails and other grossly nonnormal features of the X-X population to overcome the robust effects of its symmetry suggest that population symmetry may be one of the strongest factors favoring robustness. And the smaller robustness of the slightly asymmetric but quasi-normal Y-A population suggest that population asymmetry may be one of the factors most detrimental to robustness.

Several indices of asymmetry were explored. If a population is symmetrical, its mean and median are inevitably equal, and if it is unimodal, mean, median, and mode are all equal. Therefore the population is symmetrical about an axis through its mean, through its median or through its mode. However, for an asymmetric population, mean, median, and mode generally differ. Therefore when measuring the population's "degree" of symmetry or asymmetry, about which axis should degree of asymmetry be measured? Since we are interested in the Central Limit Effect and since, under that effect the original sampled population and all true distributions of the sample mean have a common mean, but do not necessarily have a common median or mode, the population mean seems intuitively to be the proper axis about which to measure symmetry. However, for the sake of thoroughness, symmetry about the other two axes will also be investigated.

In analogy with the nonnormality index, the index of asymmetry might be taken to be that proportion of the total population area which is not "balanced" by an equal amount of area at the exactly mirror-image location on the other side of the population mean. This would be the proportion of the population's area which did not overlap if the portion to the right of the mean were rotated 180 degrees about an axis through the mean and thus folded over upon the left portion. For any given variate-value, the amount of "nonoverlap" is simply the difference between the point probability of that variate-value and the point probability of the variate-value located an equal distance from the axis of rotation but in the opposite direction, i.e., in the case of an axis through the mean, the variate-value equal in magnitude but opposite in sign. The total amount (i.e., proportion) of nonoverlapping area is simply the aggregate of all such probability differences with the summation taking place over all variate-values having one algebraic sign. Such an index is rather crude in that it ignores the fact that in contributing to a sample mean each element of asymmetric population area is weighted by its abscissa value; and therefore in contributing to the deviation of sample mean from population mean each such element of area is weighted by its distance from its population mean. Therefore, a more refined index might be the proportion of weighted area associated with asymmetric areas,

i.e., $\dfrac{\sum_{+D} D\,|Pr(+D) - Pr(-D)|}{\sum |D\,Pr(D)|}$ where D is the distance of a variate-value from its population mean, +D represents a positive value of D, -D is the negative value of that D with the same absolute magnitude and Pr stands for "probability of". If, instead of weighting each element of area by its absolute distance from the mean, it is weighted by the square of its distance from the mean, the index of asymmetry becomes $\dfrac{\sum_{+D} D^2\,|Pr(+D) - Pr(-D)|}{\sum |D^2\,Pr(D)|}$. This s the proportion of the total population variance which is associated with asymmetric, i.e., "unbalanced", area, and is therefore an index of considerable intrinsic interest. Finally, instead of weighting by the square of the distance from the mean, one could weight by the cube of the distance so that the index of asymmetry becomes

$$\frac{\sum_{+D} D^3\,|Pr(+D) - Pr(-D)|}{\sum |D^3\,Pr(D)|}.$$

These four indices of asymmetry about an axis through the mean were quantitatively determined (primarily by the computer). So also were the analogous indices of asymmetry about the primary mode and about the median. The exact values of the modes or medians were not necessarily used. Actually the axis about which symmetry was measured was the variate-value or midpoint between variate-values which, in the former case, most nearly bisected the hump of distribution surrounding the primary mode, or which, in the latter case, lay closest to the interpolated median. Results are shown in Figures 48 to 51. For any given weighting of area, i.e., for any value of K, the relationship between the index of asymmetry and the index of nonrobustness appears to be both stronger (see figures 48, 49, and 50) and more nearly linear (see figure 51) when the axis about which symmetry is measured is the mean rather than the median or mode. (And in general the median appears to be a better axis than the mode). Thus it appears that it is asymmetry about an axis through the mean which is the really critical factor. Presumably the fairly strong relationship between nonrobustness and asymmetry about axes through the median or mode is attributable to the fact that the magnitudes of the median and mode tend to fall very close to that of the mean, and therefore asymmetry about either of the former is strongly correlated with asymmetry about the latter, which is the actually critical factor. This conclusion is supported by the fact that the relationship is fairly strong at all weights, for an axis through the mean, but, for axes through median or mode, tends to become strong only at the larger weights (see figures 48 to 50) where the area between mean and median or mode is represented by a very small proportion of the total weighted area. Finally, for an axis through the mean, the relationship between weighted areal asymmetry and nonrobustness appears to become stronger and more nearly linear (or at least monotonic) as weight increases, up to K = 2, at which point the index of asymmetry is the proportion of total population variance associated with areas which have no mirror image on the other side of an axis through the mean.

Two obvious indices of asymmetry are $\frac{\text{median - mean}}{\text{standard deviation}}$ and $\frac{\text{primary mode - mean}}{\text{standard deviation}}$, (the latter being Pearson's measure of skewness with the algebraic sign reversed). Another classical index is the "coefficient of skewness", identified as r_1 in Fisher's notation or as $\pm\sqrt{\beta_1}$ in Pearson's. In either case the coefficient of skewness is $\frac{\mu_3}{\sigma^3}$ where σ is the population standard deviation and μ_3 is the third moment about the population mean, which, in the case of a discrete distribution of variate-values, whose distance from the mean is D, is $\sum D^3\ Pr(D)$, or equivalently (since $(+D)^3\ Pr(+D) + (-D)^3\ Pr(-D) = +D^3\ [Pr(+D) - Pr(-D)]$) is $\sum_{+D} D^3[Pr(+D) - Pr(-D)]$. This index was also obtained (primarily) by the computer.

The relationship between nonrobustness and the coefficient of skewness is remarkably linear and to a lesser, but still striking, degree so is that between nonrobustness and the sigma-unit distance between median and mean. Figure 52 shows these relationships as well as those based on other measures of asymmetry and that based on nonnormality. It seems clear that nonrobustness is much more closely related to asymmetry than to nonnormality and that the relationship tends to be closest when the measure of asymmetry is sensitive to skewness rather than simply to amount of area lacking a balancing area in the exactly mirror-image location.

Turning now to the rapidity with which the Central Limit Effect produces quasi-normality in the distribution of the sample mean, the present study supports the previous one in showing that the speed is not nearly so great as commonly claimed. Graphs showing the robustness of the mean for samples from the symmetrical X-X population and from the quasi-normal Y-A population fail to show "convincingly" nonchance departures of ρ from α at any of the α values treated when sample size equals or exceeds 32, in the former case, or 256, in the latter. In all other cases the graphs appear to be showing "discernibly" real nonrobustness effects (i.e., the curves do not appear simply to be fluctuating randomly about 1.00) at some values of α at the highest sample size investigated, which was 2048 for all populations except the X population where it was 4096. As was found in the earlier study, the more extreme, i.e., the smaller, the value of α, the greater tends to be the relative departure of ρ from α at a given sample size and the larger the sample size required to reduce this departure to a prescribed value (expressed in units of size α). Thus the

RELATIONSHIP BETWEEN NONROBUSTNESS AND (WEIGHTED) AREAL ASYMMETRY ABOUT THE MEAN FOR SIX POPULATIONS.

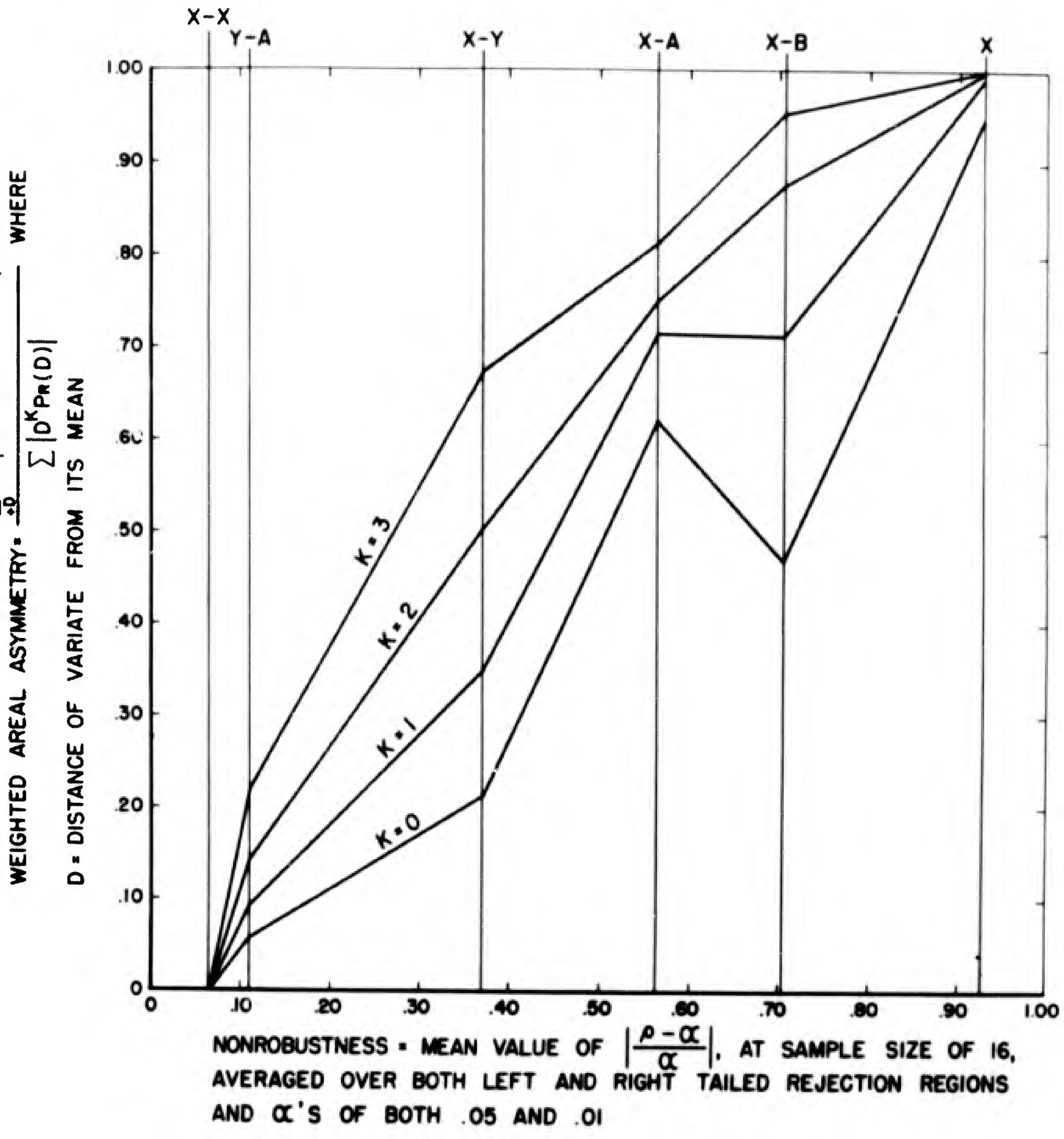

Figure 48. Influence of Population Asymmetry (about the Mean) upon the Central Limit Effect

RELATIONSHIP BETWEEN NONROBUSTNESS AND (WEIGHTED) AREAL ASYMMETRY ABOUT THE INTERPOLATED MEDIAN FOR SIX POPULATIONS

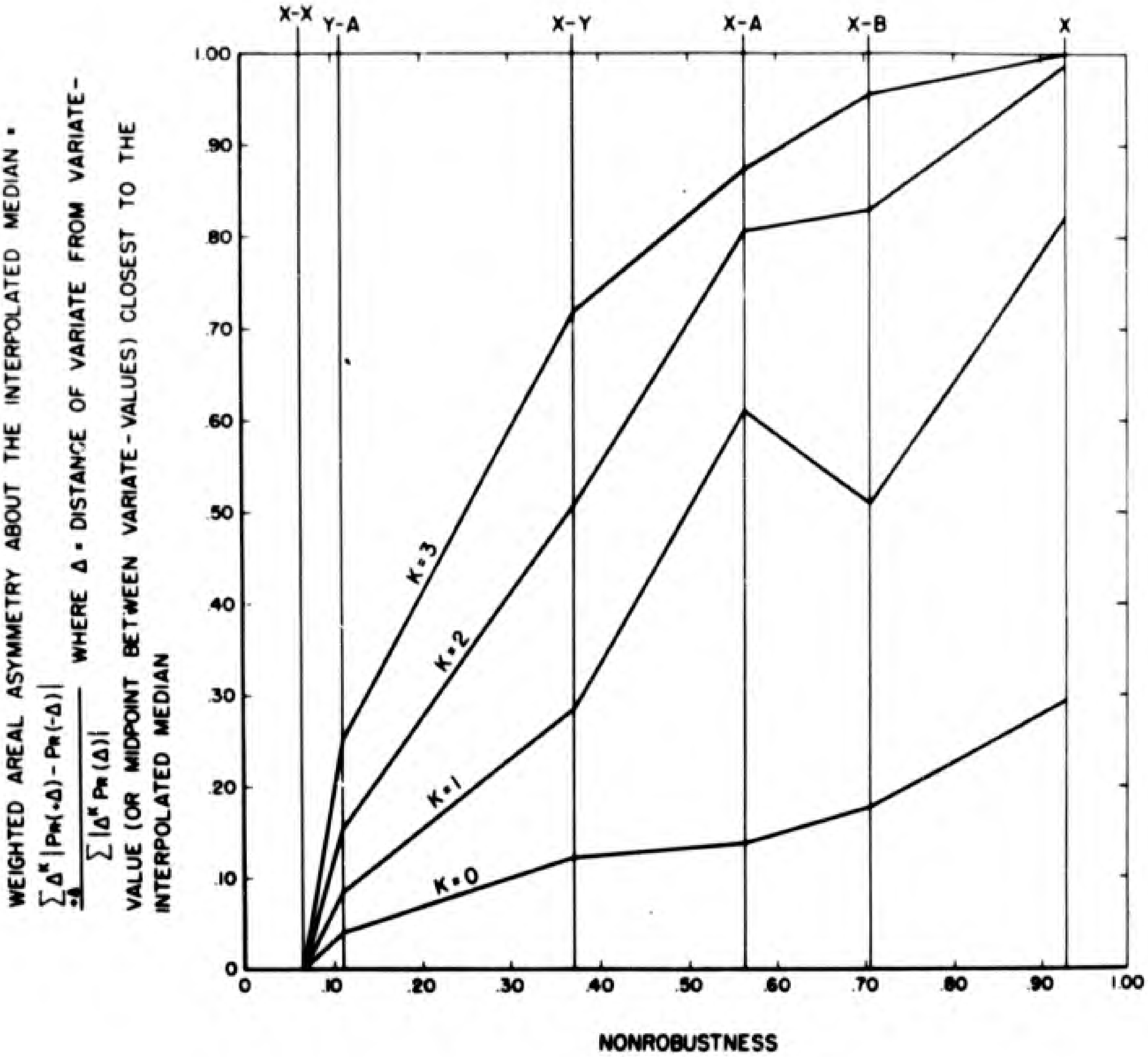

Figure 49. Influence of Population Asymmetry (about the Median) upon the Central Limit Effect

RELATIONSHIP BETWEEN NONROBUSTNESS AND (WEIGHTED) AREAL ASYMMETRY ABOUT THE MODAL AXIS FOR SIX POPULATIONS

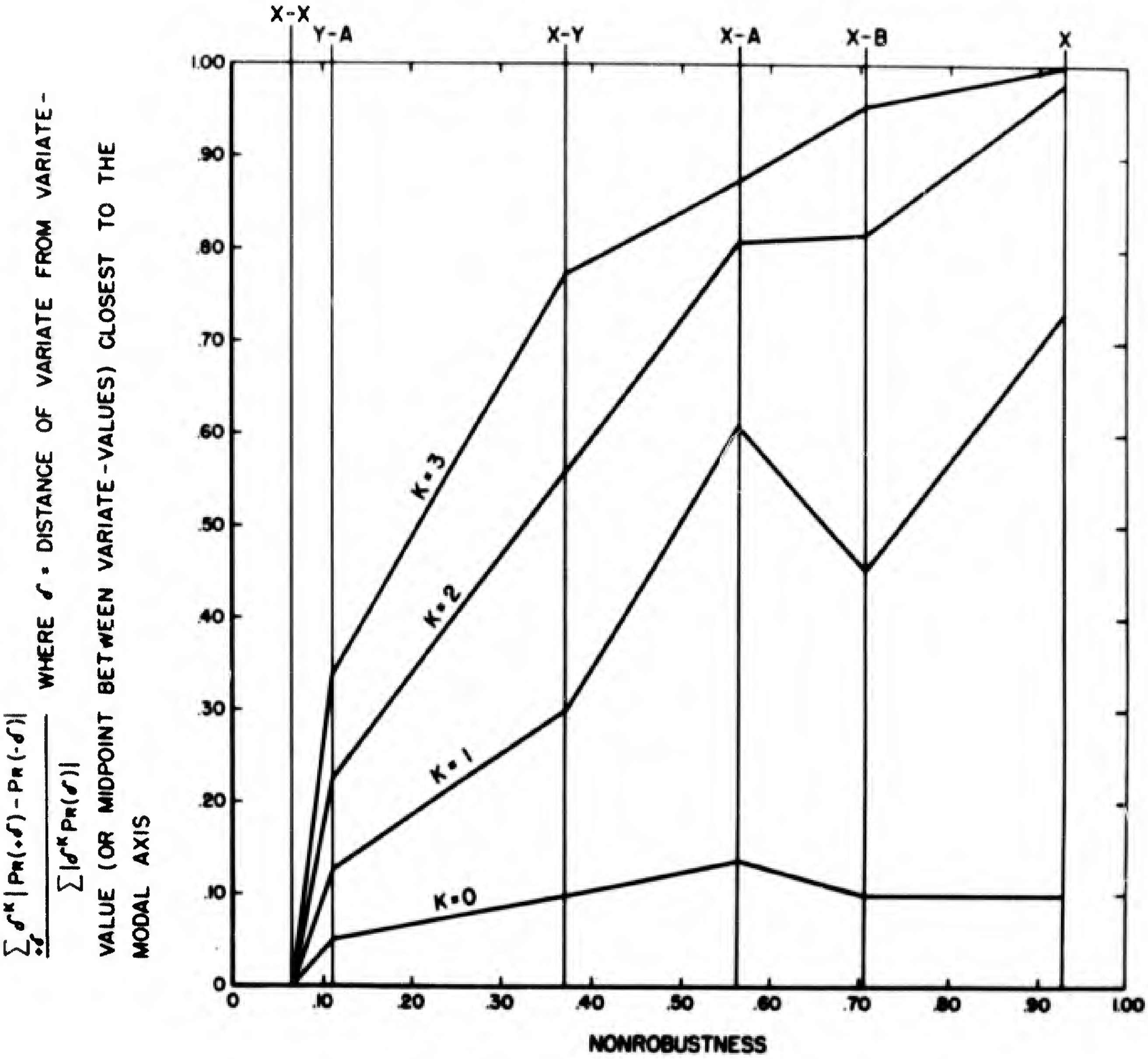

Figure 50. Influence of Population Asymmetry (about the Mode) upon the Central Limit Effect

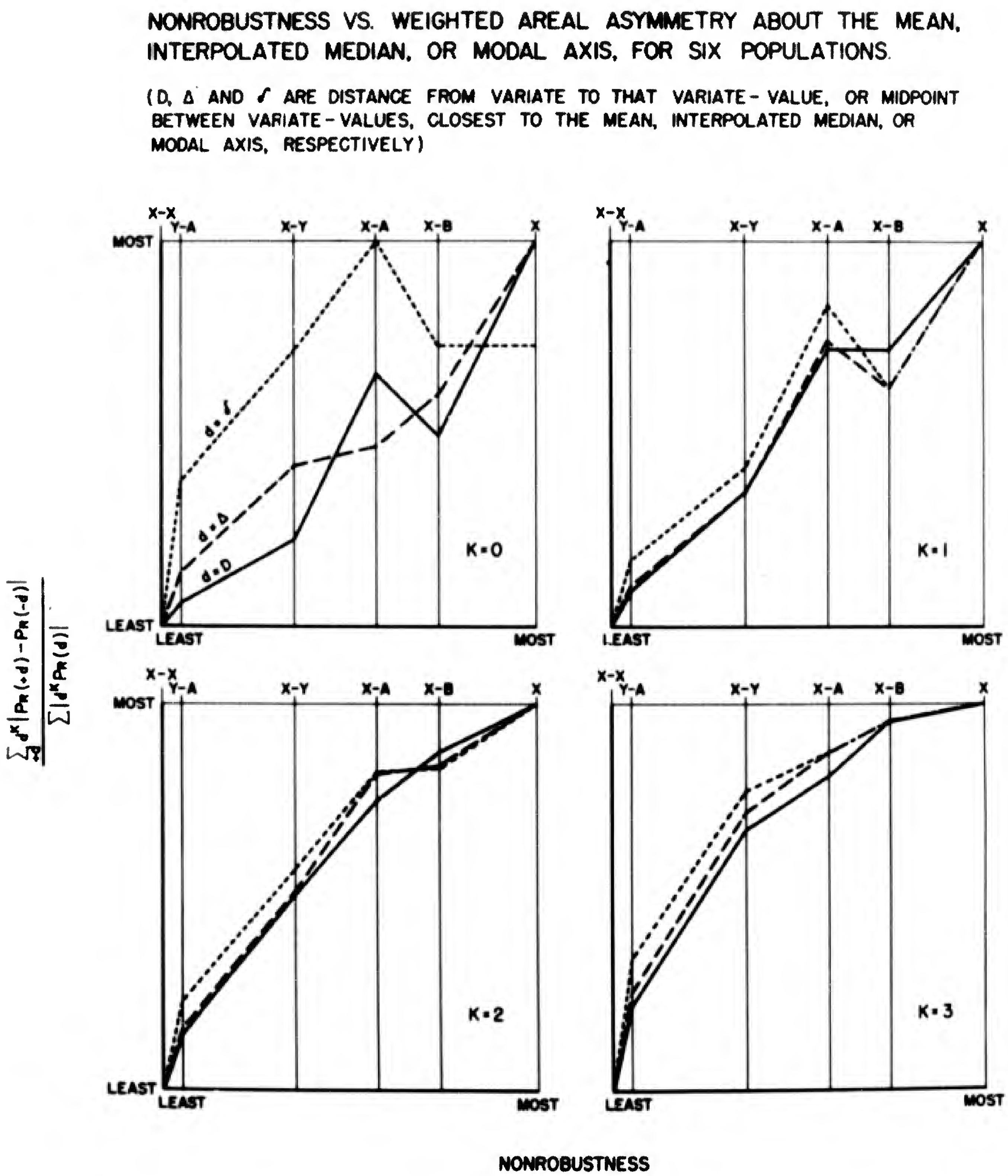

Figure 51. Dominance of Mean over Median or Mode as the Relevant Axis

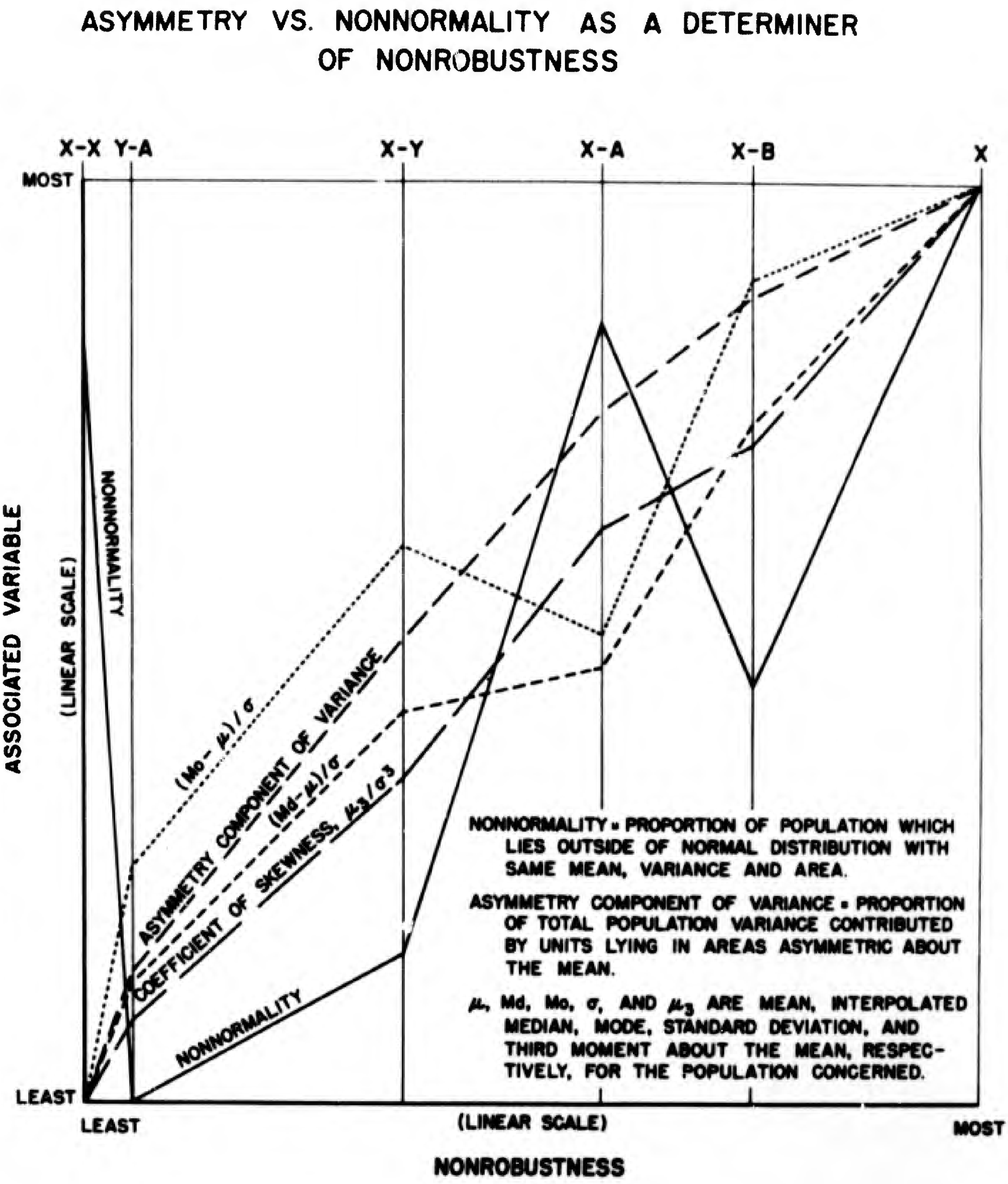

Figure 52. Dominance of Asymmetry over "Nonnormality" in Influencing Central Limit Effect

farther out one goes on a tail of the approached normal distribution, the larger the sample size required to bring the true distribution of the sample mean to within a given relative departure from the normal distribution at that point. This fact is clearly apparent in all of the data. For $\alpha \geq .20$ the ratio of ρ/α or α/ρ tends to be fairly close to 1.00 at all but the smallest sample sizes, and for all of the cases investigated. However, for α's $\leq .10$, i.e., for α's representing the "testing tails", the curves representing the ρ/α or α/ρ ratio tend to "fan out", roughly speaking, about a pivot point whose ordinate is 1.00 and whose abscissa is usually a very large value, perhaps exceeding the largest sample size investigated. The smaller α and the smaller the sample size, the greater tends to be the departure of the curve from 1.00. Thus the central portion of the distribution of the sample mean tends to be extremely "robust", under all but the very worst conditions, while the portion at the testing tails tends to be rather "nonrobust" under all but the most favorable conditions. In a sense, therefore, the Central Limit Effect produces conditions more favorable to estimation than to hypothesis testing.

II Robustness of One-Sample t Tests

The one-sample TX test was investigated and thoroughly discussed in the earlier study. The present study simply adds data which extend results up to $N = 4096$ and make results more statistically reliable at $N \leq 1024$. The graph based upon the combination of all data, both old and new, is especially interesting, however, in that it demonstrates that even when sample size exceeds 4000, at the testing tails the true significance levels are still in a state of rather steep approach to their limiting values of α (and may still be distantly separated from α).

The TD 2N statistic has more in common with one-sample than with two-sample tests and is therefore discussed here. The shape of the population of correlated difference-scores from which its observations were drawn is unknown. However the extent of the correlation can be inferred from data presented in appendix IV of the earlier report. The TD 2N statistic takes account of an existing correlation which the otherwise analogous TXY 2N2N statistic ignores. A comparison of the graphs for these two statistics shows TD 2N to be somewhat, but not spectacularly, more robust than TXY 2N2N.

III Robustness of Two-Sample Z Tests

In the earlier report the writer attempted to explain in detail the statistical factors operating to produce the results obtained for each statistic investigated. Results of the present study can be rationalized in much the same way; however, owing to the very large number of statistics investigated, the writer will not attempt it. Instead, he will list the general types of factors which need to be considered in order properly to understand why the results came out as they did. The results are, in fact, consonant with statistical logic as the dedicated reader may discover for himself.

Consider first the two-sample Z statistic. Its numerator, in the null case, is the difference between two sample means, and its denominator is the true value of the standard error of this difference between means. The denominator is in fact the true standard error of the difference between sample means irrespective of whether or not the sampled populations are normally distributed. Furthermore, no assumption of homogeneity of variance is required, since the denominator takes explicit account of any heterogeneity of variance by using the true values of each population variance in calculating the true standard error of the numerator. Thus since there is no assumption of homogeneity to violate and the denominator is unaffected by nonnormality, we must look to the numerator for an explanation of the robustness or nonrobustness of Z. Normal theory assumes that the observations from which each mean was calculated were drawn from a normal population. If this is the case then each sample mean will be normally distributed and the difference between the two sample means will also be normally distributed, and, consequently, the Z statistic will be the ratio of a normally distributed variate, i.e., the difference between means, to the standard deviation of that variate. Thus, if we wish, we may discard the "twice removed" assumption of normally distributed populations and substitute for it either the equivalent and "once removed" assumption of normally distributed sample means, or the equally equivalent assumption of a normally distributed difference between means. The latter substitute assumption is the more basic; however, the assumption of normally distributed means is perhaps more conducive to insight and will be temporarily employed for that reason after which we will revert to the more basic assumption.

By the Central Limit Theorem (or, to be more precise, by the empirical evidence for it, based on finite-sized samples) the larger the number of observations upon which a mean is based, the more nearly normal becomes the true sampling distribution of that mean. So if all other factors are held constant, the two-sample Z statistic comes closer to meeting its assumptions (as rephrased) as the size of either sample is increased, (provided of course that the population from which that sample was drawn is a nonnormal one). And as a consequence of more nearly satisfied assumptions one might expect an increase in robustness.

There is another important factor to be considered, however. The extent to which the shape of the distribution of one mean can influence the shape of the distribution of the difference between two means depends upon the relative variances of the distributions of the two means involved. For example, if one mean has a true distribution which is normal with a variance of 100 and the other mean has a true distribution which is nonnormal but with a variance of 1, the small variance of the nonnormally distributed mean will prevent that nonnormality from having much effect upon the distribution of the difference between the two means. Contrariwise if the true distributions of both means have the same variance, the nonnormality will have greater influence, and if the original situation is reversed so that the nonnormal distribution has the larger variance, then its nonnormality should greatly influence the shape of the true distribution of the difference between the two means.

Thus the larger the absolute sample size, the "less nonnormal" becomes the shape, and the smaller becomes the variance, σ^2/N, of the true distribution of the sample mean. So the effect of increasing the size of one sample while holding constant the size of the other sample is to increase the conformity of the first sample's mean with normal theory while diminishing its influence upon the robustness of the test. However there is no guarantee that the reduction in the increased sample's variance will not vitiate whatever benefit might otherwise have accrued from the increased robustness of its mean. If the increased sample's mean was originally more robust than that of the sample held constant, it is impossible to say whether the robustness of Z will be increased by the increased robustness of the increased sample's mean or decreased by its diminished weight in compensating for the nonrobustness of the nonincreased sample's mean. Hence, increased robustness of the components of the Z statistic does not necessarily insure increased robustness for the Z statistic itself.

The situation is further complicated by a third factor, interaction between (the effects of) shapes of the distributions of the two sample means. While only a single nonnormal shape is to be found among the four populations (X, A, Y, and B) from which observations were drawn, a multitude of nonnormal shapes, a different shape for each sample size, are to be found in the distributions of the means of samples of observations drawn from either of the populations, X or A, having that single nonnormal shape. If the size of one sample is increased while that of the other sample is held constant (or even increased nonproportionately) the individual distributions of the means of the two samples can only become more nearly normal (or remain equally normal or nonnormal). However, this does not guarantee that the shape of the distribution of the really critical variable, the difference between the two means, will only change in such a way as to make $\bar{X}_1 - \bar{X}_2$ more robust. This topic will be expanded upon later.

Thus the robustness of the two-sample Z statistic may be regarded to be a simultaneous function of (a) the extent to which the absolute size of a sample has brought about quasi-normality in the distribution of its mean (for those samples drawn from nonnormal populations), (b) the extent to which the relative variances of the two sampling distributions of the two sample means weight the shapes of those respective distributions in influencing the shape of the distribution of the difference between the two means, and (c) the interaction between the shapes of the distributions of the two sample means. To the extent that the interaction effect does not operate in opposition to those of the other two factors, or to the extent that it is relatively small, robustness effects can be largely explained in terms of factors (a) and (b). Therefore, since interaction effects are difficult to predict, factor (c) will be temporarily ignored in the discussion to follow.

Table IV gives the variance of the true distribution of the mean for both samples, as well as the ratio of these two variances, for each Z statistic investigated. It therefore provides explicit information relevant to (b), above, and should facilitate understanding of the robustness effects obtained for the two-sample Z statistic. Consider the sequence of ZPQ statistics formed by arranging them in order of increasing relative sizes of the P sample: ZPQ N3N,

TABLE IV

RELATIONSHIP BETWEEN VARIANCES OF TRUE DISTRIBUTIONS OF SAMPLE MEANS

	Variance of the Mean for		Variance of the Mean in Units of σ^2/N for		
ZPQ C_1NC_2N	Sample from P Population	Sample from Q Population	Sample from P Population	Sample from Q Population	Ratio
ZXX NN	$\frac{\sigma^2}{N}$	$\frac{\sigma^2}{N}$	1	1	1
ZXX N3N ZXY N3N	$\frac{\sigma^2}{N}$	$\frac{\sigma^2}{3N}$	1	$\frac{1}{3}$	3
ZXX N2N ZXY N2N	$\frac{\sigma^2}{N}$	$\frac{\sigma^2}{2N}$	1	$\frac{1}{2}$	2
ZXX 2N2N ZXY 2N2N	$\frac{\sigma^2}{2N}$	$\frac{\sigma^2}{2N}$	$\frac{1}{2}$	$\frac{1}{2}$	1
ZXX 2NN ZXY 2NN	$\frac{\sigma^2}{2N}$	$\frac{\sigma^2}{N}$	$\frac{1}{2}$	1	$\frac{1}{2}$
ZXX 3NN ZXY 3NN	$\frac{\sigma^2}{3N}$	$\frac{\sigma^2}{N}$	$\frac{1}{3}$	1	$\frac{1}{3}$
ZXA N3N ZXB N3N ZYA N3N	$\frac{\sigma^2}{N}$	$\frac{\sigma^2}{12N}$	1	$\frac{1}{12}$	12
ZXA N2N ZXB N2N ZYA N2N	$\frac{\sigma^2}{N}$	$\frac{\sigma^2}{8N}$	1	$\frac{1}{8}$	8
ZXA 2N2N ZXB 2N2N ZYA 2N2N	$\frac{\sigma^2}{2N}$	$\frac{\sigma^2}{8N}$	$\frac{1}{2}$	$\frac{1}{8}$	4
ZXA 2NN ZXB 2NN ZYA 2NN	$\frac{\sigma^2}{2N}$	$\frac{\sigma^2}{4N}$	$\frac{1}{2}$	$\frac{1}{4}$	2
ZXA 3NN ZXB 3NN ZYA 3NN	$\frac{\sigma^2}{3N}$	$\frac{\sigma^2}{4N}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{4}{3}$

ZPQ N2N, ZPQ 2N2N, ZPQ 2NN, ZPQ 3NN. In this sequence (i) the absolute size of the P sample either increases or remains constant, (ii) that of the Q sample either decreases or remains constant, and (iii) the relative size of the P sample increases and therefore the relative variance of its mean diminishes as does the relative influence of its mean upon the distribution of ZPQ. Now suppose that P is one of the nonnormal populations, X or A, and Q is one of the "normal" populations, Y or B. Then, if it changes at all, the robustness of the P sample's mean must increase, because of (i) while its influence upon Z must diminish, because of (iii). The robustness of the Q sample's mean will be complete, and therefore constant and unaffected by (ii), at all sample sizes (if discreteness effects at small N's can be ignored), and its influence upon Z will increase, because of (iii). Thus, the P sample's contribution becomes more robust and receives smaller weight relative to the Q sample's contribution which is "perfectly" robust (except perhaps around N = 2) in all cases. So both factors (a) and (b) operate in the direction of greater robustness. ZXY and ZXB meet the stated conditions and their robustness increases in the sequence listed. "ZAY" would meet the conditions and ZYA increases in robustness in the sequence ZYA 3NN, ZYA 2NN, ZYA 2N2N, ZYA N2N, ZYA N3N. Thus predictions made on the basis of factors (a) and (b) alone are consonant with the results obtained for ZXY, ZXB, and ZYA. In the case of ZXA and ZXX, where both samples violate the normality assumption, prediction, even on the basis of factors (a) and (b) alone, is not always possible. However, when the respective sample sizes change from N and 2N to 2N and 2N, an originally less robust sample is increasing from less to equal robustness while decreasing in influence. On the basis of factors (a) and (b) alone, one would predict increased robustness for Z. This is supported by the results for ZXX but refuted by those for ZXA 2NN and ZXA 2N2N. When sample sizes for ZXX change from N and N to N and 2N, a sample originally equal in both robustness and influence to those of a second sample increases in robustness while diminishing in influence. On the basis of factors (a) and (b) alone, one would predict an increase in robustness for Z, but the opposite appears to have occurred. Presumably factor (c) has manifested its influence in these cases where prediction failed.

Ignoring factor (c) amounts to regarding the robustness of Z as the weighted sum of the individual robustnesses of $\overline{X}_1$ and $\overline{X}_2$, when it is actually the robustness of $\bar{X}_1 - \bar{X}_2$, the difference between means (regarded as a single variable with a single distribution). By replacing the substitute assumption of normally distributed sample means with the equivalent, but more basic, substitute assumption of a normally distributed difference between means, all three factors (a), (b), and (c) are subsumed under the single factor of shape of the distribution of the difference between means. However, while prediction now has a more valid basis, it has also become more difficult insofar as the effect of relative sample size is concerned. By way of compensation, the effect of absolute sample size becomes more easily predictable.

Consider the statistic

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

in the case investigated in this report where N_2 is some integral multiple, C, of N_1 (or vice versa). Its robustness may be regarded as determined by a single factor, the shape of the distribution of its numerator. If we can establish that the Central Limit Theorem applies to its numerator as absolute sample size increases while relative sample size remains constant, then the increasing robustness of Z is established for these same conditions. If the numerator can be expressed as the mean of a single variable, however complex so long as its distribution has a finite variance, then the Central Limit Theorem does apply to the numerator. This will be accomplished in the following derivation in which we let $N_1 = N$ and $N_2 = CN$,

$$\bar{X}_1 - \bar{X}_2 = \frac{1}{N} \sum_1^N X_1 - \frac{1}{CN} \sum_1^{CN} X_2$$

$$= \frac{1}{N} \sum_1^N X_1 - \frac{1}{CN} \left[\sum_1^N X_{2i} + \sum_{N+1}^{2N} X_{2i} + \ldots + \sum_{(C-1)N+1}^{CN} X_{2i} \right]$$

or, letting $j = N + i$, $k = 2N + i$, etc.

$$= \frac{1}{N}\sum_{1}^{N} X_1 - \frac{1}{CN}\sum_{1}^{N}\left[X_{2i} + X_{2j} + X_{2k}\text{-----}\right]$$

$$= \frac{1}{N}\sum_{1}^{N}\left[X_1 - \frac{1}{C}X_{2i} - \frac{1}{C}X_{2j} - \frac{1}{C}X_{2k} - \text{-----}\right]$$

= mean of N values of the variable $X_1 - \frac{1}{C}X_{2i} - \frac{1}{C}X_{2j} - \frac{1}{C}X_{2k}$ -----

= mean of N values of the variable $X_1 - \frac{1}{C}\sum_{1}^{C} X_2$.

Thus the numerator of ZXA N2N is the mean of a variable $X - \frac{1}{2}A' - \frac{1}{2}A''$ where X, $\frac{1}{2}A'$ and $\frac{1}{2}A''$ are component variables, the last two of which have the same distribution (i.e., are two different random selections of a single variable, $\frac{1}{2}A$). The Central Limit Theorem therefore holds, the numerator of two-sample Z becomes more nearly normally distributed as N increases while relative sample size remains constant, and consequently so does Z itself. Hence the general robustness of all of the Z statistics investigated in this study should increase (presumably monotonically) with increasing values of N. By general robustness is meant the overall fit between the true and normal theory sampling distributions of Z. However improved overall fit does not necessarily imply improved fit at every point and the approach of $\bar{\rho}$ toward α with increasing N need not be monotonic as a glance at the curve labeled $\alpha = .20$ for ZX will show. Despite the fact that monotonicity of approach of $\bar{\rho}$ to α is not guaranteed, however, it (or something very close to it) does appear always to have been the case at the testing tails.

The decrease in the robustness of ZXX as sample sizes change from N and N to N and 2N is now understandable in the light of the following facts which have now been established: (i) the robustness of all of the Z statistics investigated in this study depends exclusively upon the Central Limit Effect (as it applies to the entire numerator of Z, treated as a single variable), (ii) the Central Limit Effect appears to be highly sensitive to asymmetry, becoming rapidly attenuated by increasing degrees of asymmetry, (iii) the single variable, X - X, of which the numerator of ZXX NN is the mean, is symmetrically distributed. However the single variable, $\frac{1}{2}X + \frac{1}{2}X - X$, of which the numerator of ZXX 2NN is the mean is not symmetrically distributed, since exchanging or permuting the actual numerical values of the three X's does not generally result in values of $\frac{1}{2}X + \frac{1}{2}X - X$ which are equal in absolute magnitude but opposite in sign.

The variable of which the numerator of Z is the mean takes one of the forms

$$X_S - \frac{1}{C}\sum_1^C X_L$$

or

$$\frac{1}{C}\sum_1^C X_L - X_S$$

where X_S and X_L stand for the variable of which the smaller and the larger samples, respectively, were taken. These two forms of the variable may be replaced by $\bar{X}_S - \bar{X}_L$ and $\bar{X}_L - \bar{X}_S$ where it is understood that $\bar{X}_S$ is the "mean" of a single value of X_S and $\bar{X}_L$ is the mean of C values of X_L. Now suppose that the two original sampled populations are asymmetric and identical. Then, if sample sizes are equal, i.e., if $C = 1$, the distributions of $\bar{X}_S$ and $\bar{X}_L$ are also identical. Since they are identical, the probability of drawing an observation of value a from the $\bar{X}_S$ population and an observation of value b from the $\bar{X}_L$ population is the same as the probability of drawing the a from the $\bar{X}_L$ population and the b from the $\bar{X}_S$ population. Therefore the probability that $\bar{X}_S - \bar{X}_L = a - b$ and the probability that $\bar{X}_S - \bar{X}_L = b - a = -(a - b)$ are the same. Since, for any values of a and b, the observations $a - b$ and $-(a - b)$ have equal probabilities, the $\bar{X}_S - \bar{X}_L$ distribution is symmetrical about zero. And, as we already know, this situation is highly conducive to robustness of Z. Now suppose that C assumes values in the increasing sequence, 2, 3, 4, ---- ∞. By the Central Limit Theorem, the shape of the $\bar{X}_L$ population becomes increasingly less nonnormal and therefore increasingly dissimilar to the shape of the nonnormal $\bar{X}_S$ population with which it was originally identical. Likewise the variance of the $\bar{X}_L$ distribution diminishes monotonically from the variance of the $\bar{X}_S$ distribution with which it was originally identical. The disparity in both shape (provided that $\bar{X}_S$ is quite nonnormal) and variance between the two populations will be quite large even at $C = 2$ and will increase at a decelerating rate as C increases until, at C = infinity, $\bar{X}_L$ has zero variance about a value of μ and can be replaced by the constant μ in the expression $\bar{X}_S - \bar{X}_L$, the variance and shape of whose distribution consequently become the variance and shape of the distribution of $\bar{X}_S$, i.e., of X. Thus as relative sample size departs from equality to a maximum of inequality (as defined by the condition that $C = \infty$) the shape of the distribution of $\bar{X}_S - \bar{X}_L$ changes from exact symmetry toward a limiting shape which is identical to that of the original asymmetric sampled population. To the extent that the increasing dissimilarity between the $\bar{X}_S$ and $\bar{X}_L$ distributions with increasing dissimilarity in sample size causes increasing asymmetry in the distribution of $\bar{X}_S - \bar{X}_L$ (as one would suspect generally to be the case), one would (on the basis of empirical evidence already presented) expect increasing nonrobustness in the Z statistic. Thus as the size of one sample is held constant while that of a second sample from the same asymmetrical population is increased, one would expect the robustness of Z to diminish, very nearly monotonically, at a decelerating rate which may be very rapid at the beginning if the size of the sample held constant is small. One would therefore expect the robustness of ZXX to diminish as sample sizes change from N, N to N, 2N to N, 3N, and the results confirm this expectation.

As C increases, $\bar{X}_L$ becomes more nearly normal, and the variance of its distribution diminishes. The A population is identical to the X population except that it has smaller variance. The Y population has the same variance as X but is normally distributed. And the B population is normal with a smaller variance than the X population. Therefore when $\bar{X}_S = X$ and $\bar{X}_L$ is the mean of C values of A, Y or B, the distribution of $\bar{X}_L$, and therefore of $\bar{X}_S - \bar{X}_L$, may be regarded as simply "anticipating" part of the effect which would have been produced by an increase in C if X_L had also equaled X. To the extent that this is true, the preceding argument for ZXX applies also to ZXA, ZXY, and ZXB, provided that it is the X sample whose size is held constant while the size of the other sample is increased. One would therefore expect their robustness to diminish as sample sizes change from N, 2N to N, 3N and it does (although only very slightly for ZXB).

The robustness of ZXX diminished as sample sizes changed from N, N to 2N, N to 3N, N, that of ZXY and ZXA diminished as sample sizes changed from N, 2N to N, 3N and that of ZYA diminished from 2N, N to 3N, N; and, finally, that of ZXY, ZXA, and ZXB diminished when sample sizes changed from 2N, N to 2N, 2N. In all these cases robustness diminished as one sample size was held constant while the other sample increased in size. Thus robustness diminished as a consequence of taking more data! And therefore robustness may sometimes be increased by discarding data! Another interesting paradox lies in the fact that ZXX 2N2N is far more robust than ZXY 2N2N. Thus in this case robustness is greater when both populations violate the normality assumption in the same way (i.e., have identical nonnormal shapes) than when one population violates the normality assumption in that way but the other population, in effect, satisfies the assumption. Thus, in a sense, robustness is greater when the assumption is fully violated than when it is "half" violated!

To summarize, the robustness of all of the Z statistics investigated in this study can be regarded as determined exclusively by the Central Limit Effect operating upon the mean of a sample of N observations from the distribution of a single variable. Consequently, the <u>general</u> robustness of all Z statistics investigated should increase monotonically with increasing N, i.e., the overall fit between the true and normal-theory distributions of Z should improve monotonically with increasing N. (This does not guarantee that the approach of $\bar{\rho}$ to α will be monotonic for a single specified α.) Furthermore, as N increases the true significance level estimated by ρ should approach a limiting value of α. In the two-sample case the single variable mentioned above is a composite variable the shape of whose distribution is considerably influenced by relative sample size and the relative shapes and variances of the sampled populations. Therefore, although the two-sample Z test requires only the single assumption of normality, when that assumption is violated the robustness of the test is considerably influenced by (a) the absolute sizes of the samples, (b) the relative sizes of the samples, (c) the absolute shapes of the sampled populations, especially the extent to which both are symmetric, (d) the relative shapes of the sampled populations, especially the "extent" to which they are identical, and (e) the relative variances of the sampled populations.

The Z statistics enjoy the advantage of requiring a single assumption, the violation of which affects only the numerator of a ratio. Furthermore, in all cases investigated here, the numerator can be reduced to a single variable subject to the Central Limit Effect. As a consequence of these features (a) certain qualitative aspects of the robustness of Z, such as the general increase in robustness with increasing sample size, are fairly predictable on an <u>a priori</u>

basis, (b) certain regularities appear in the behavior of Z's robustness, such as the virtually uniform tendency for ρ to approach α monotonically at the testing tails or the tendency for robustness to either diminish or increase in the sample size sequence 3NN, 2NN, 2N2N, N2N, N3N for all Z statistics except ZXX for which robustness diminished according to the last half of the sequence. However, while the trend of Z's robustness is relatively predictable and well behaved, the extent of its robustness is quite variable ranging from the close conformity of ZYA N3N to normal theory to the violent departure of ZXB N3N from it. Therefore while the unqualified claim that the two-sample Z test is robust is refuted by the data, these same data indicate that it may be relatively robust under certain conditions and that, more important, knowing those conditions, that relative robustness may be fairly reliably predicted on an *a priori* basis.

IV Robustness of Two-Sample t Tests

Turning now to the two-sample t statistic, it is necessary to consider the following facts in order properly to understand why the results came out as they did. (a) Since analogous Z and t statistics have identical numerators and since the denominator of a given Z statistic is a constant, what has been said and shown about the robustness of Z is equally relevant to the "robustness" of the numerator of the corresponding t statistic. (b) There is a positive correlation between mean and variance of samples drawn from the X population (see appendix III of the earlier report) and this must be equally true of samples drawn from the A population. Therefore, there is a positive correlation between $\bar{X}$ and $\sum(X-\bar{X})^2$ and between $\bar{A}$ and $\sum(A-\bar{A})^2$, i.e., between one of the variable terms in the numerator and one of the variable terms in the denominator of any t statistic based on one or more samples from either the X or the A population. (c) When sampling from the X population (or, equivalently, from the A population) the sample variance is extremely nonrobust, assuming extreme values, both large and small, far more frequently than would be expected under normal theory. (This is shown in figure 53 on page 110.) Consequently, the sums of squares in the denominator of t can be expected to assume, at far greater than normal-theory "frequency", values which are extremely large or extremely small relative to their normal-theory expectations. (d) The two variable terms in the denominator of t, $\sum(X_1-\bar{X}_1)^2$ and $\sum(X_2-\bar{X}_2)^2$, may be written as $(N_1-1)\ \hat{\sigma}^2_{X_1}$ and $(N_2-1)\ \hat{\sigma}^2_{X_2}$. Therefore, in contributing to the squared denominator of t, the sample estimates of their population variances are weighted by one less than the sizes of their respective samples. (Alternatively, one can replace a sum of squares by the product of sample variance and sample size; therefore, in contributing to the squared denominator of t, the sample variances are weighted by their sample sizes.) Thus any distortion from normal-theory in the distribution of a sample variance will, in effect, influence the robustness of the squared denominator of t in nearly direct proportion to the relative size of the sample. (e) When the assumption of homogeneity of variance is violated and samples are of unequal size, the squared denominator of the two-sample t statistic cannot be expected to remain an unbiased estimate of the variance of the true distribution of the numerator. In the case of the one-sample t statistic (which involves a single population variance and a single sample size) the squared denominator always provided an unbiased estimate of the variance of the true distribution of the numerator. Furthermore, as sample size increased the denominator varied less and less about the parameter it estimated, and this variance shrank at a rate greater than that with which the variance of the numerator diminished with increasing sample size. (See table 29, page 209, of ref. 4 and table 1 of ref. 5.) The result was that as sample size increased the variability of the numerator contributed more and more to the variability of t, relative to the contribution by the variability of the denominator. Since the squared denominator was an unbiased estimate of the normal-theory value required (and since the numerator was - in relative terms - fairly "robust") the distribution of the one-sample t approached its normal-theory distribution as sample size increased. (This argument is documented in appendix III of the earlier report.) In the case of the two-sample t statistic, the sample estimates of their individual population variances are still unbiased, but those estimates are, in effect, combined to produce an estimate of the variance of the true distribution of the difference between the two sample means, and this latter estimate is unbiased when variances are homogeneous or when samples are of equal size but generally biased otherwise. These points can be shown algebraically as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sum (X_1-\bar{X}_1)^2 + \sum (X_2-\bar{X}_2)^2}{N_1 + N_2 - 2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

$$= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(N_1-1)\,\hat{\sigma}^2_{X_1} + (N_2-1)\,\hat{\sigma}^2_{X_2}}{N_1 + N_2 - 2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

$$= \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(N_1-1)\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}{N_1-1 + N_2-1}\hat{\sigma}^2_{X_1} + \frac{(N_2-1)\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}{N_1-1 + N_2-1}\hat{\sigma}^2_{X_2}}}$$

Now, since $\hat{\sigma}^2_{X_1}$ and $\hat{\sigma}^2_{X_2}$ are unbiased estimates, the true means of their sampling distributions are, by definition, $\sigma^2_{X_1}$ and $\sigma^2_{X_2}$, respectively. Therefore by the rules of linear combination of variables, the entire expression under the radical sign is a variable the mean of whose true sampling distribution is simply the same expression with $\sigma^2_{X_1}$ and $\sigma^2_{X_2}$ substituted for $\hat{\sigma}^2_{X_1}$ and $\hat{\sigma}^2_{X_2}$ respectively. Therefore, again by definition, the expression containing $\hat{\sigma}^2_{X_1}$ and $\hat{\sigma}^2_{X_2}$ is an unbiased estimate of the same expression with $\sigma^2_{X_1}$ and $\sigma^2_{X_2}$ substituted for their estimates. We will now show that when these substitutions are made the denominator reduces to $\sigma_{\bar{X}_1-\bar{X}_2}$, i.e., the standard deviation of the true distribution of the numerator, provided that population variances are homogeneous or sample sizes are equal, but not generally otherwise. First, substituting the population parameters for their estimates, we obtain

$$\sqrt{\frac{(N_1-1)\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}{N_1-1 + N_2-1}\sigma^2_{X_1} + \frac{(N_2-1)\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}{N_1-1 + N_2-1}\sigma^2_{X_2}}$$

then letting $\sigma^2_{X_1} = \sigma^2_{X_2} = \sigma^2$, which introduces the "assumption" of homogeneity of variance, we have

$$\sqrt{\frac{(N_1-1)\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}{N_1-1+N_2-1}\sigma^2 + \frac{(N_2-1)\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}{N_1-1+N_2-1}\sigma^2}$$

$$= \sqrt{\frac{(N_1-1)\left(\frac{1}{N_1}+\frac{1}{N_2}\right)+(N_2-1)\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}{N_1-1+N_2-1}\sigma^2} = \sqrt{\frac{(N_1-1+N_2-1)\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}{N_1-1+N_2-1}\sigma^2}$$

$$= \sqrt{\left(\frac{1}{N_1}+\frac{1}{N_2}\right)\sigma^2} = \sqrt{\frac{\sigma^2}{N_1}+\frac{\sigma^2}{N_2}} = \sqrt{\frac{\sigma^2_{X_1}}{N_1}+\frac{\sigma^2_{X_2}}{N_2}} = \sqrt{\sigma^2_{\bar{X}_1}+\sigma^2_{\bar{X}_2}}$$

which, since the variance of the difference between two uncorrelated variates is equal to the sum of their separate variances,

$$= \sqrt{\sigma^2_{\bar{X}_1-\bar{X}_2}} = \sigma_{\bar{X}_1-\bar{X}_2}$$

Now, returning to the radical with which we started and letting $N_1 = N_2 = N$, we will show that if sample sizes are equal the same result can be derived, i.e., unbiassedness can be proved, without making any assumption of homogeneity of variance.

$$\sqrt{\frac{(N-1)\left(\frac{1}{N}+\frac{1}{N}\right)}{N-1+N-1}\sigma^2_{X_1} + \frac{(N-1)\left(\frac{1}{N}+\frac{1}{N}\right)}{N-1+N-1}\sigma^2_{X_2}} = \sqrt{\frac{(N-1)\left(\frac{2}{N}\right)}{2(N-1)}\left(\sigma^2_{X_1}+\sigma^2_{X_2}\right)}$$

$$= \sqrt{\frac{1}{N}\left(\sigma^2_{X_1}+\sigma^2_{X_2}\right)} = \sqrt{\frac{\sigma^2_{X_1}}{N}+\frac{\sigma^2_{X_2}}{N}} = \sqrt{\sigma^2_{\bar{X}_1}+\sigma^2_{\bar{X}_2}} = \sqrt{\sigma^2_{\bar{X}_1-\bar{X}_2}} = \sigma_{\bar{X}_1-\bar{X}_2}$$

Since $\sigma_{\bar{X}_1-\bar{X}_2}$, or the equivalent $\sqrt{\frac{\sigma^2_{X_1}}{N_1}+\frac{\sigma^2_{X_2}}{N_2}}$, is the denominator of the two-sample Z statistic (which has the same numerator as the corresponding two-sample T statistic), we have, in effect, "derived" Z from T by substituting the two population variances for their sample estimates and by making "assumptions" appropriate to the normal-theory t test. This is not surprising, since the t test was designed to perform the function of the Z test when population variances are unknown and therefore must be estimated from the samples. Each of the derivations presented achieved its results through a clearing of terms. Letting $\sigma^2_{X_1} = \sigma^2_{X_2} = \sigma^2$ permitted a combining of terms involving N_1-1 and N_2-1 into a single coefficient for σ^2, which then reduced to $\left(\frac{1}{N_1} + \frac{1}{N_2}\right)$. Letting $N_1 = N_2 = N$ permitted the combination of N_1 and N_2 into terms which cancelled. However, these two cases exhaust the possibilities involving a clearing of terms and are therefore the only general cases in which the squared denominator of t is an unbiased estimate of the variance of the numerator. Individual cases may arise, when variances are heterogeneous and sample sizes are unequal, in which, by extreme good luck, $\sigma^2_{X_1}$, $\sigma^2_{X_2}$, N_1 and N_2 have values such that

$$\sqrt{\frac{(N_1-1)\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}{N_1+N_2-2}\sigma^2_{X_1} + \frac{(N_2-1)\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}{N_1+N_2-2}\sigma^2_{X_2}} = \sigma_{\bar{X}_1-\bar{X}_2}$$

However, while the possibility exists, it is scarcely of even academic interest since one does not know the values of $\sigma^2_{X_1}$ and $\sigma^2_{X_2}$ when he uses the t test and he is therefore ill equipped to solve the above formula and thereby learn of his possible good fortune.

To summarize, factors which can influence the robustness of the two-sample T statistics investigated in this study are: (a) all factors which influence the robustness of the corresponding Z statistic, which may include both absolute and relative shapes of the sampled populations, relative variances of the sampled populations, and both absolute and relative sample sizes (which can be reduced to the absolute and relative shapes and the relative variances of the true distributions of the two sample means, or reduced still further to the shape of the distribution of the single variable of which the numerator is the mean of a sample of N values), (b) a positive correlation between $\bar{X}$ and $\sum(X-\bar{X})^2$ and, equally, between $\bar{A}$ and $\sum(A-\bar{A})^2$, (c) the extreme nonrobustness of $\sum(X-\bar{X})^2$, and, equally, of $\sum(A-\bar{A})^2$, (d) the fact that since the variable terms in the denominator of t can be written $\sum(X_1-\bar{X}_1)^2 = N_1 s_1^2 = (N_1-1)\,\hat{\sigma}^2_{X_1}$ and $\sum(X_2-\bar{X}_2)^2 = N_2 s_2^2 = (N_2-1)\,\hat{\sigma}^2_{X_2}$, the robustness or nonrobustness of the true distribution of each sample's variance influences the robustness or nonrobustness of the squared denominator of t in very nearly direct proportion to the size of the sample to which it belongs, (e) the fact that the squared denominator of t is an unbiased estimate of the desired parameter, $\sigma^2_{\bar{X}_1-\bar{X}_2}$, when population variances are homogeneous or when samples are equal in size, but not generally otherwise. When only the assumption of normality is violated, the factors listed under (a), (b), (c), and (d) above, will influence the robustness of the two-sample T statistic; when only the assumption of homogeneity of variance is violated the factor listed under (e) will influence robustness (as will other factors - it is not implied that it is only the mean of the true sampling distribution of the denominator which is distorted by heterogeneity of variance).

It is clear from the above that a multitude of factors may be brought into play when the normality assumption is violated. For a fixed ratio of sample sizes, as N increases any distortions from the normal-theory t distribution attributable to (a) should diminish due to the Central Limit Effect, while a rapidly diminishing variability of the denominator relative to that of the numerator should attenuate the effects of factors (b), (c), and (d), although the attenuated effect need not diminish monotonically. The various factors may tend to cause distortions, i.e., deviations of ρ from α, in opposite directions. Furthermore, the distortions which they individually cause may diminish at different rates as N increases with the result that different factors are prepotent at different values of N in producing the net distortion attributable to a combination of factors. Thus, ρ may greatly exceed α at certain values of N and be greatly exceeded by α at others, with the result that the graphed curves appear far on one side of 1.00, crossing this ordinate very precipitously and assuming values far on the other side of 1.00 as N increases. The situation is far simpler when only the assumption of homogeneity of variance is violated. In that case, only factor (e), of those listed, is involved. However, while the other factors individually produce distortions which should diminish (although perhaps not always monotonically) as N increases, the distortion attributable to (e) is in the nature of a bias and therefore does not diminish toward a distortion of zero but rather migrates toward the degree of distortion characteristic of the true value of the bias. Therefore, one would not expect the deviation of ρ from α to approach zero as N increased, even at very large values of N. Consequently one would not expect the ratios graphed to approach an ordinate of 1.00 even at the highest values of N, and indeed they do not. They do, however, in the case of simple heterogeneity, tend to remain on a single side of 1.00, which is what would be expected.

The effect of violation of the assumption of homogeneity alone is shown by the YB graphs. The graph for TYB 2N2N shows curves whose deviations from an ordinate of 1.00 are systematic and convincingly "real" at $N \leq 4$ but which are visually indistinguishable from chance at $N \geq 32$; in all cases they are small relative to the deviations obtained at a corresponding N value in other graphs for T or Z. The graphs for TYB 2NN, TYB 3NN, TYB N2N, and TYB N3N show curves which above $N = 4$ are practically straight horizontal lines, running parallel to but far removed from the line through an ordinate of 1.00. In general, the distance separating a given curve from the line through an ordinate of 1.00 is greater when sample sizes are 3N and N than when they are 2N and N. Both effects are explainable in terms of a biased denominator the bias of which increases with the relative discrepancy between sample sizes.

The effect of violation of the single assumption of normality is shown in the TXX graphs (for identical nonnormal populations) and in the TXY graphs (for one nonnormal and one normal population). Some of these graphs support the contention that violation of the single assumption of normality permits each of a variety of factors to produce distortions from normal-theory, each in its own idiosyncratic way, the distortions varying perhaps in direction, extent and speed of reduction with increasing N. The curves in all cases appear to approach a limiting ordinate of 1.00, but the approach is sometimes quite definitely non-monotonic and this includes cases of TXX as well as TXY. Some of the graphs showed curves which extended far on either side of 1.00.

The TXA, TXB, and TYA graphs show the effect of adding violation of the assumption of homogeneous variances to one of the assumption-violating situations already considered and graphed. For example, TXB 2NN (in a sense) shows the effect of heterogeneity of variance in addition to, and in interaction with, the effect of the particular violation of the normality assumption represented by TXY 2NN. In fact, one can consider TXB 2NN to embody the violations inherent in TXY 2NN and TYB 2NN. However, this perspective is a rather misleading one since it implicitly interprets the normality assumption as applying to each of the two sampled populations when actually the assumption applies basically (and therefore more validly) to the single distribution of the difference between means. In no case where sample sizes are unequal and the assumption of homogeneity of variance is violated does any curve approach a limiting value of 1.00. And frequently the curves will lie far above the horizontal line through 1.00 at certain values of N and far below it at other N values. In such cases the curve tends to pass through 1.00 in steep ascent or descent. Consequently, just above the N at which the curve crosses the line through 1.00, robustness decreases with increasing absolute sample size and we are presented with yet another paradox, that robustness can sometimes be increased by using a smaller absolute sample size!

For cases in which population variances were heterogeneous, the overall robustness (i.e., robustness assessed over all ten N values, with special regard for the highest N's) of the appropriate T statistic diminished with increasing disparity of sample sizes, i.e., diminished as sample sizes changed from 2N and 2N, to 2N and N, to 3N and N, or from 2N and 2N, to N and 2N, to N and 3N. This is about what one would expect on the basis of our previous considerations. However, it again presents the intriguing result that robustness can <u>increase</u> as a result of discarding data, e.g., by throwing away the last N observations in the second sample, TYA N3N is converted to the more robust TYA N2N.

When the "overall" robustness of a T statistic is compared with that of the corresponding Z statistic, it is found that the Z statistic is more robust in all (investigated) cases involving heterogeneous variances, irrespective of whether sample sizes are equal or unequal. (This "overall" superiority does not necessarily extend to every N value of every graph, however.) When variances are homogeneous the T statistic is sometimes the more robust. In comparisons of TXX 3NN vs ZXX 3NN, TXX 2NN vs ZXX 2NN, TXY 3NN vs ZXY 3NN, and TXY 2NN vs ZXY 2NN, the T statistic is the more robust in all left-tail cases, most two-tail cases, and in one right-tail case. In all other cases the Z test was superior or equal. The evidence therefore is not conducive to a concise generalization concerning the robustness of Z relative to T when variances are homogeneous. There is, of course, an element of subjectivity in comparisons of "overall" robustness, and the reader is invited to make these comparisons for himself.

The data admit of many more meaningful comparisons than the types mentioned. However, the possibility will simply be mentioned, and the reader may explore for himself such comparisons as interest him. The author is not particularly interested in cataloging the myriad specific constellations of conditions under which a test is "robust" or "nonrobust". He _is_ interested in pointing out five facts. First, even under a liberal definition of robustness the two-sample t test is simply not very robust, or to put it more accurately, the test is drastically nonrobust under many of the conditions investigated in this study and relatively robust under few conditions. Therefore the unqualified claim that the two-sample t test is robust simply does not stand up under the evidence. Second, when population variances are heterogeneous and samples are unequal in size, the distribution of the two-sample t (with perhaps certain rare exceptions of academic interest only) does not approach the normal-theory t distribution as its limiting distribution at N = infinity. Thus even a very liberal criterion of robustness may _never_ be met at _any_ sample sizes if population variances and sample sizes are sufficiently unequal. (A ratio of 2/1 for both sample sizes and population standard deviations was sufficient to produce drastic departures of p from α even at N = 1024 in the present study.) Thirdly, the clear non-monotonicity with which some curves did approach 1.00 and, particularly, the passage of other curves from far above 1.00, across 1.00 in steep descent, to far below 1.00 (or vice versa) as N increased clearly establish the fact that the robustness of two-sample t does not necessarily improve as N increases, at all values of N, but may worsen drastically, instead. Fourth, robustness may diminish as sample sizes become more nearly equal and it may diminish as a result of taking more data. Fifth, the two-sample t test is neither universally inferior nor universally superior to the corresponding two-sample Z test.

Turning now to the R statistics, i.e., the ratios of two sample variances, they are extremely nonrobust and were expected to be. The earlier study showed that $\chi^2 = \frac{\sum (X-\bar{X})^2}{\sigma^2}$ is violently nonrobust (see figure 53). Therefore, since

ROBUSTNESS OF THE SAMPLE VARIANCE FOR SAMPLES CONSISTING OF N OBSERVATIONS DRAWN FROM THE X POPULATION

RATIO BETWEEN EMPIRICAL, ρ, AND NORMAL-THEORY, α, ONE-TAILED CUMULATIVE PROBABILITIES FOR VALUES OF $S^2 = \frac{\Sigma_1^N (X-\bar{X})^2}{N}$ (OR, EQUIVALENTLY, OF $\hat{\sigma}^2 = \frac{\Sigma_1^N (X-\bar{X})^2}{N-1}$ OR $\chi^2 = \frac{\Sigma_1^N (X-\bar{X})^2}{\sigma^2}$) HAVING NORMAL-THEORY CUMULATIVE PROBABILITY OF α

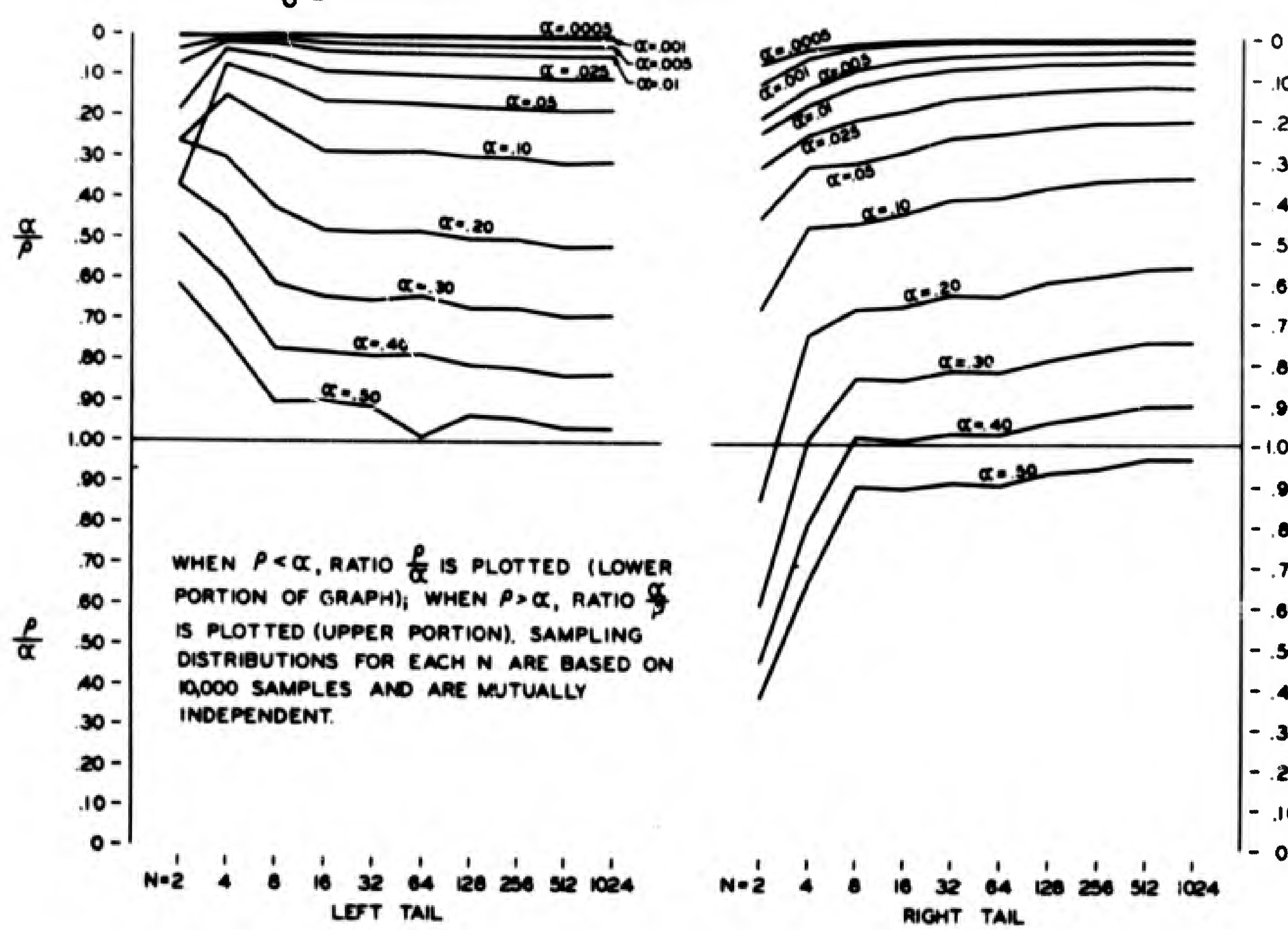

Figure 53. Robustness of the Sample Variance, i.e., of the Ingredients from which the R Statistic is Obtained. (Data are from previous study; figure is based on Tables 20 and 21 of reference 4.)

$$\text{RXX 2N2N} = \frac{\dfrac{\sum (X_1-\bar{X}_1)^2}{2N-1}}{\dfrac{\sum (X_2-\bar{X}_2)^2}{2N-1}} = \frac{\sum (X_1-\bar{X}_1)^2}{\sum (X_2-\bar{X}_2)^2} = \frac{\sigma^2 \chi^2_{x_1}}{\sigma^2 \chi^2_{x_2}} = \frac{\chi^2_{x_1}}{\chi^2_{x_2}} \quad \text{and}$$

$$\text{RXY 2N2N} = \frac{\sum (X-\bar{X})^2}{\sum (Y-\bar{Y})^2} = \frac{\chi^2_x}{\chi^2_y} \quad ,$$

and since the two chi squares in the ratio are independent of one another, it would be reasonable to expect the R statistics to be drastically nonrobust. As was the case for χ^2_x in the earlier study, the curves for the R statistics do not approach 1.00 as an asymptote (indeed at early N values the curves for RXY 2N2N depart increasingly from 1.00 as N increases); instead they tend to level out at values far removed from 1.00.

VI Robustness of the Analysis-of-Variance F Test

When samples are of equal size, and all assumptions are met, the F statistic can be written

$$F = \frac{\dfrac{N \sum_1^k (\bar{X}_i-\bar{\bar{X}})^2}{k-1}}{\dfrac{\sum_1^k \sum_1^{N_i} (X_{ij}-\bar{X}_i)^2}{k(N-1)}} = \frac{N\, \hat{\sigma}_{\bar{X}}^2}{\dfrac{\sum_1^k \hat{\sigma}_X^2}{k}}$$

Since the variance estimates are unbiassed, the means of their true distributions are the population variances which they estimate. Therefore, letting n and d stand for numerator and denominator respectively, and letting E( ) and Var( ) denote the mean, i.e., expected value, and variance respectively of the true distribution of the statistic enclosed within the parentheses,

$$E(n) = N\, \sigma_{\bar{X}}^2 = \frac{N\, \sigma_X^2}{N} = \sigma_X^2$$

$$E(d) = \frac{1}{k} \sum_1^k \sigma_X^2 = \frac{k}{k}\, \sigma_X^2 = \sigma_X^2 = E(n)$$

and, since the variance of a variance estimate, $\hat{\sigma}_X^2 = \sum_1^N (X-\bar{X})^2/(N-1)$, based on normally distributed X's from the same population (see ref. 6, pp. 151 and 135), is $Var(\hat{\sigma}_X^2) = 2\sigma_X^4/(N-1)$, we have, under normal theory,

$$Var(n) = Var(N\hat{\sigma}_{\bar{X}}^2) = N^2 Var(\hat{\sigma}_{\bar{X}}^2) =$$

$$\frac{2N^2 \sigma_{\bar{X}}^4}{k-1} = \frac{2N^2}{k-1}\left(\frac{\sigma_X^4}{N^2}\right) = \frac{2\sigma_X^4}{k-1}$$

$$Var(d) = Var\left(\frac{1}{k}\sum_1^k \hat{\sigma}_X^2\right) = \frac{1}{k^2}\sum_1^k Var(\hat{\sigma}_X^2)$$

$$= \left(\frac{k}{k^2}\right)\left(\frac{2\sigma_X^4}{N-1}\right) = \frac{2\sigma_X^4}{k(N-1)}$$

For a fixed number of treatments, k, the variance of the denominator is almost inversely proportional to sample size, N, while the variance of the numerator does not depend upon N and is a constant. Therefore, as N approaches infinity, the variance of the denominator approaches zero while that of the numerator remains fixed. Thus when N becomes infinite, the variation of the denominator about its expected value becomes zero and the denominator becomes simply E(d). (See appendix VII.) Therefore,

at $N = \infty$, $E(F) = E\left(\frac{n}{d}\right) = E\left[\frac{n}{E(d)}\right] = \frac{E(n)}{E(d)} = \frac{\sigma^2}{\sigma^2} = 1$ and

$$Var(F) = Var\left(\frac{n}{d}\right) = Var\left[\frac{n}{E(d)}\right] = \frac{Var(n)}{[E(d)]^2} = \frac{Var(n)}{[E(n)]^2}$$

This ultimate variance of F therefore depends upon the variance of the numerator, but not upon that of the denominator. It is influenced by the location, but not the shape, of the denominator's distribution. The speed with which this condition is approached can be roughly inferred from

$$\frac{Var(d)}{Var(n)} = \frac{\frac{2\sigma^4}{k(N-1)}}{\frac{2\sigma^4}{k-1}} = \frac{k-1}{k(N-1)} = \frac{(k-1)/k}{(N-1)}$$

The (k-1)/k term is capable of contributing relatively little to the ratio, having a minimum value of 1/2 when k = 2 and a maximum of 1 when k is infinite. Therefore, it matters little whether k be regarded as a constant or as a variable. The fact that Var (d)/Var (n) approaches zero as N approaches infinity indicates that Var (F) is becoming more and more determined by Var (n) relative to the influence of Var (d). (The relationship between the variances of a ratio, its numerator, and its denominator, is not a simple one; nevertheless, Var (d)/Var (n) appears to be a serviceable measure of the dwindling influence of the denominator's variance upon the variance of F.) Figure 54 plots this

EFFECT OF SAMPLE SIZE UPON RELATIVE VARIANCES OF DENOMINATOR, d, AND NUMERATOR, n, OF $F = \dfrac{\dfrac{\Sigma_i^k N_i(\bar{X}_i - \bar{X})^2}{k-1}}{\dfrac{\Sigma_i^k \Sigma_i^{N_i}(X_{ij} - \bar{X}_i)^2}{\Sigma_i^k (N_i - 1)}}$ WHEN ALL $N_i = N$ AND ALL ASSUMPTIONS ARE MET

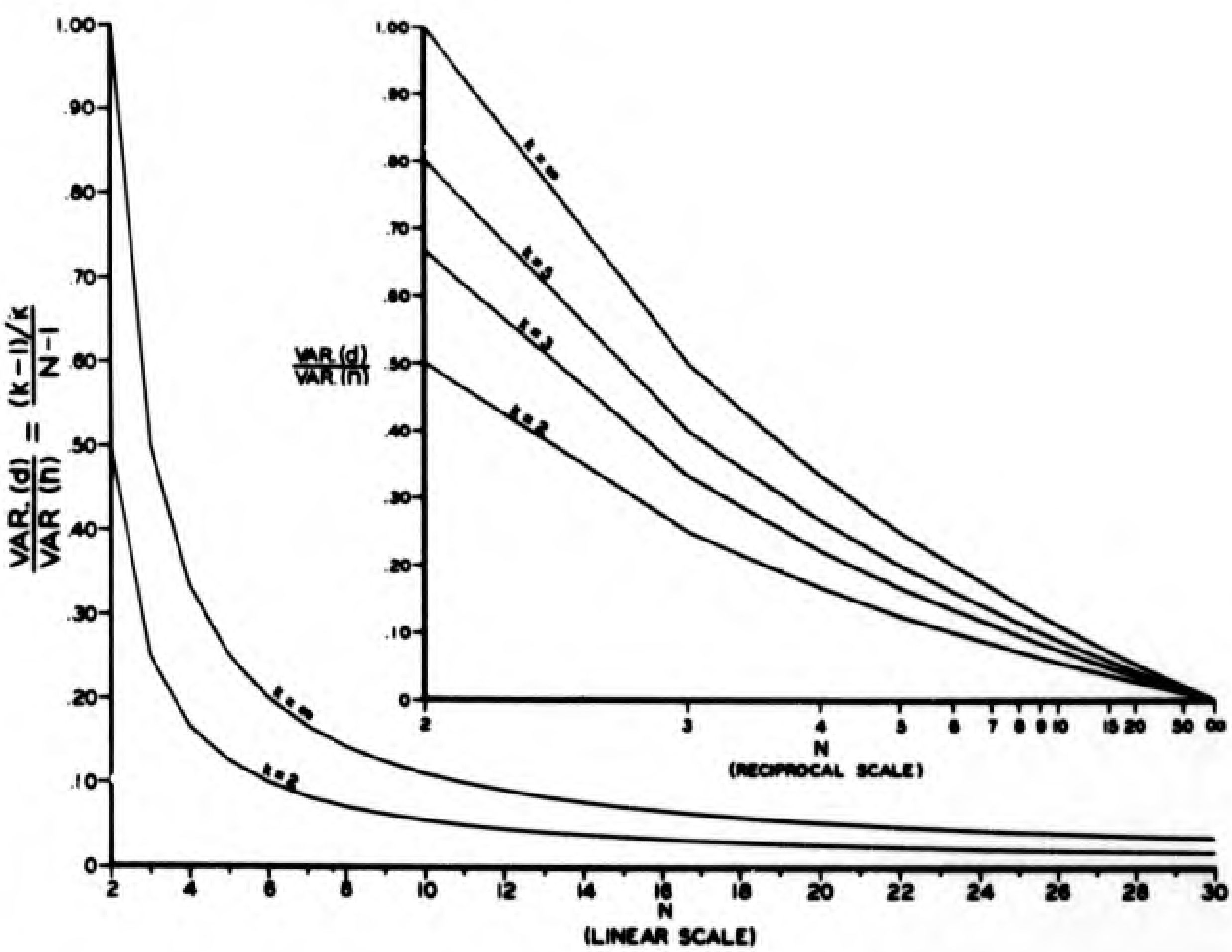

Figure 54. Diminishing Variability of Denominator, Relative to Numerator, of F as Sample Size Increases

ratio as ordinate against sample size as abscissa, both for linear and reciprocal abscissa scales. The figure treats only the normal-theory case where all assumptions are met. Under nonnormality or heterogeneity of variance, the rate at which the ratio approaches zero may be affected but, as will presently be shown, the ultimate value of the ratio, at infinite N, is still zero.

The above results were derived for the case where all $N_i = N$. This restriction was unnecessary and was employed only because of the obviousness of the derivation it permitted. Under normal theory

$$F = \frac{\dfrac{\sum_1^k N_i (\bar{X}_i - \bar{\bar{X}})^2}{k-1}}{\dfrac{\sum_1^k \sum_1^{N_i} (X_{ij} - \bar{X}_i)^2}{\sum_1^k (N_i - 1)}} = \frac{\dfrac{\chi^2_{k-1} \sigma_X^2}{k-1}}{\dfrac{\chi^2_{\sum_1^k (N_i-1)} \sigma_X^2}{\sum_1^k (N_i-1)}} = \frac{n}{d}$$

and since the mean and variance of a chi-square variate are its degrees of freedom and twice its degrees of freedom, respectively, all of the previous results follow immediately, except that $\sum_1^k (N_i - 1)$ must be substituted for $k(N-1)$ in the expressions involving Var (d).

The paragraphs immediately to follow are concerned with the ultimate robustness of F, i.e., its robustness when the size of its smallest sample is infinite. In order for F to be completely robust under assumption-violation when $N_{min} = \infty$, its ultimate distribution would have to be identical to the ultimate distribution of normal-theory F. Two necessary conditions for this identity are that

$$\text{Ult E (F)} = \text{Ult Normal-Theory E (F)} \quad \text{and}$$
$$\text{Ult Var (F)} = \text{Ult Normal-Theory Var (F)}$$

It will be shown that irrespective of nonnormality and/or heterogeneity of variance,

as $N \to \infty$

$\text{Var (d)} \to 0$ and

$\dfrac{\text{Var (d)}}{\text{Var (n)}} \to 0$ so that (See appendix VII.)

$F = \dfrac{n}{d} \to \dfrac{n}{E\ (d)}$ and

$\text{Var (F)} \to \dfrac{\text{Var (n)}}{[E\ (d)]^2}$

BLANK PAGE

and that, irrespective of nonnormality, so long as variances are homogeneous,

$$\text{Ult E (n)} = \text{Ult E (d)} = \sigma^2 = \text{Ult Normal-Theory Value, so that}$$

$$\text{Ult E (F)} = E\left[\frac{n}{E\ (d)}\right] = \frac{E\ (n)}{E\ (d)} = 1 = \text{Ult Normal-Theory E (F) and that}$$

$$\text{Ult Var (n)} = \text{Ult Normal-Theory Var (n) so that}$$

$$\text{Ult Var (F)} = \frac{\text{Var n}}{[E\ (d)]^2} = \frac{\text{Normal Theory Var (n)}}{[\text{Normal Theory E (d)}]^2} = \text{Ult Normal-Theory Var (F)}$$

and therefore that F is ultimately completely robust under nonnormality alone, and that when variances are heterogeneous (regardless of the presence or absence of nonnormality),

$$\text{Ult E (F)} \neq \text{Ult Normal-Theory E (F), except when all } N_i = N$$

$$\text{Ult Var (F)} \neq \text{Ult Normal-Theory Var (F), except when } k = 2 \text{ and } N_1 = N_2$$

so that under heterogeneity of variance multi-sample F is not ultimately completely robust and the situation tends to worsen with increasing disparity in sample sizes. The degree of ultimate nonrobustness under heterogeneity of variance will be quantified for two important classes of cases, and F will be shown to be highly sensitive to fairly small degrees of heterogeneity under otherwise realistic sampling conditions even when sample size is infinite.

The derivations consist mainly of algebraic manipulations upon the working formula for F (i.e., F calculated from sample observations) which, in and of itself assumes nothing about the sampled populations. A few additional formulae are introduced into the algebraic manipulations, but they either apply to the general case, and therefore assume neither normality nor homogeneity, or they assume normality of the distribution of the sample mean and are introduced in the case where N is infinite so that the validity of the assumption is guaranteed by the Central Limit Theorem. The derivations therefore apply to the general case (free of assumptions concerning normality or homogeneity) except where explicitly stated restrictions limit their application to a more specific case. The derivations provide the mathematical means of taking account of nonnormality and heterogeneity, should they occur, but neither their occurrence nor their nonoccurrence is assumed until explicitly introduced in interpreting the obtained results.

The proofs will make use of the facts (see refs. 6 and 19) that (a) if $L = C_1V_1 + C_2V_2 + \cdots + C_nV_n$, where the C's are constants and the V's are arbitrary variates, and if the true mean and variance of V_i are $\overline{V}_i$ and σ_i^2, respectively, then the true mean of L is $\sum_{i=1}^{n} C_i\overline{V}_i$ and, if the V_i are independent, the true variance of L is $\sum_{i=1}^{n} C_i^2\ \sigma_i^2$, (b) the variance of $\hat{\sigma}^2 = \sum_i^n (X-\overline{X})^2/(n-1)$ is $\frac{\phi - \sigma^4}{n} + \frac{2\sigma^4}{n(n-1)}$ for any population for which a fourth moment, ϕ, about the mean exists, (c) if X is a normally distributed variate with variance σ^2, then $\sum_1^n (X-\overline{X})^2 = \chi^2_{n-1}\ \sigma^2$ where χ^2_{n-1} is chi-square with n-1 degrees of freedom, (d) if χ_f^2 is a chi-square variate with f degrees of freedom, its true mean is f and its true variance is 2f, (e) the sum of several independent chi-squares is distributed as chi square with degrees of freedom equal to the sum of the individual degrees of freedom of the chi-squares contributing to the sum.

Let there be C distinguishable populations all having the same mean, μ, and variances σ_i^2 (i=1, 2, --- C) which may or may not differ. Let K_i be the number of samples, all of the <u>same</u> size N_i, drawn from the ith population. (This restriction is not very severe since an additional distinguishable population can be "created" for each different sample size.) Let $K = \sum_{i=1}^{c} K_i$ and $T = \sum_{i=1}^{c} K_i\ N_i$. Let P_{ijk} be the kth observation in the jth sample from the ith population, $\overline{P}_{ij}$ be the mean of the N_i observations in the jth sample drawn from the ith population, $\overline{\overline{P}}_i$ be the mean of all K_i samples, each of size N_i, drawn from the ith population, and G be the grand mean of all observations drawn. Then instead of summing over all K samples at once, the summing will be done in two stages by partitioning the $K = \sum_1^c K_i$ samples into C sets each containing a varying number, K_i, of samples, i.e., by replacing $\sum_1^K$ by the entirely equivalent $\sum_1^c \sum_1^{K_i}$. Then,

$$F = \frac{\dfrac{\sum_1^k N_i (\bar{X}_i - \bar{\bar{X}})^2}{k-1}}{\dfrac{\sum_1^k \sum_1^{N_i} (X_{ij} - \bar{X}_i)^2}{\sum_1^k (N_i - 1)}} = \frac{\dfrac{\sum_{i=1}^c \sum_{j=1}^{K_i} N_i (\bar{P}_{ij} - G)^2}{K-1}}{\dfrac{\sum_{i=1}^c \sum_{j=1}^{K_i} \sum_{k=1}^{N_i} (P_{ijk} - \bar{P}_{ij})^2}{\sum_1^c \sum_1^{K_i} (N_i - 1)}} = \frac{n}{d}$$

Before proceeding further, certain equivalences, which will be needed later, will be developed.

$$G = \frac{\sum_{i=1}^c \sum_{j=1}^{K_i} \sum_{k=1}^{N_i} P_{ijk}}{\sum_{i=1}^c \sum_{j=1}^{K_i} N_i} = \frac{\sum_{i=1}^c \sum_{j=1}^{K_i} N_i \bar{P}_{ij}}{\sum_{i=1}^c K_i N_i} = \frac{\sum_{i=1}^c K_i N_i \bar{\bar{P}}_i}{T}$$

$$\text{Var } (G) = \sigma_G^2 = \frac{\sum_{i=1}^c K_i^2 N_i^2 \sigma_{\bar{\bar{P}}_i}^2}{T^2} = \frac{\sum_{i=1} K_i N_i \sigma_i^2}{T^2}$$

$$E (\bar{P}_{ij} - \mu)^2 = E [(\bar{P}_{ij} - \mu) - o]^2 = E [(\bar{P}_{ij} - \mu) - E (\bar{P}_{ij} - \mu)]^2$$

$$= \text{Var } (\bar{P}_{ij} - \mu) = \text{Var } (\bar{P}_{ij}) = \sigma_{\bar{P}_{ij}}^2$$

$E (\mu - G)^2 = \sigma_G^2$ by analogous derivation

$$\sum_{j=1}^{K_i} (\bar{P}_{ij} - G)^2 = \sum_{j=1}^{K_i} [(\bar{P}_{ij} - \bar{\bar{P}}_i) + (\bar{\bar{P}}_i - G)]^2$$

$$= \sum_{j=1}^{K_i} (\bar{P}_{ij} - \bar{\bar{P}}_i)^2 + 2(\bar{\bar{P}}_i - G)(K_i \bar{\bar{P}}_i - K_i \bar{\bar{P}}_i) + K_i(\bar{\bar{P}}_i - G)^2$$

$$= \sum_{j=1}^{K_i} (\bar{P}_{ij} - \bar{\bar{P}}_i)^2 + K_i(\bar{\bar{P}}_i - G)^2$$

$$\sum_{i=1}^{c} N_i \sum_{j=1}^{K_i} (\bar{P}_{ij}-G)^2 = \sum_{i=1}^{c} N_i \sum_{j=1}^{K_i} [(\bar{P}_{ij}-\mu) + (\mu-G)]^2$$

$$= \sum_{i=1}^{c} N_i \sum_{j=1}^{K_i} (\bar{P}_{ij}-\mu)^2 + 2(\mu-G) \sum_{i=1}^{c} (N_i K_i \bar{\bar{P}}_i - N_i K_i \mu)$$

$$+ (\mu-G)^2 \sum_{i=1}^{c} N_i K_i$$

$$= \sum_{i=1}^{c} N_i \sum_{j=1}^{K_i} (\bar{P}_{ij}-\mu)^2 + 2(\mu-G)(TG-T\mu) + T(\mu-G)^2$$

$$= \sum_{i=1}^{c} N_i \sum_{j=1}^{K_i} (\bar{P}_{ij}-\mu)^2 - T(\mu-G)^2$$

Proceeding now to show that the variance of the denominator shrinks toward zero, both absolutely and relative to the variance of the numerator, as sample size approaches infinity:

$$F = \frac{\dfrac{\sum_{i=1}^{c} N_i \sum_{j=1}^{K_i} (\bar{P}_{ij}-\bar{\bar{P}}_i)^2 + \sum_{i=1}^{c} N_i K_i (\bar{\bar{P}}_i-G)^2}{K-1}}{\dfrac{\sum_{i=1}^{c} \sum_{j=1}^{K_i} \sum_{k=1}^{N_i} (P_{ijk}-\bar{P}_{ij})^2}{\sum_{i=1}^{c} \sum_{j=1}^{K_i} (N_i-1)}} = \frac{n}{d}$$

$$d = \frac{\sum_{i=1}^{c} \sum_{j=1}^{K_i} (N_i-1)\, \hat{\sigma}_{ij}^2}{\sum_{i=1}^{c} K_i (N_i-1)}$$

$$\text{Var}(d) = \frac{\sum_{i=1}^{c} \sum_{j=1}^{K_i} (N_i-1)^2 \, \text{Var}(\hat{\sigma}_{ij}^2)}{\left[\sum_{i=1}^{c} K_i (N_i-1)\right]^2}$$

$$= \frac{\sum_{i=1}^{c} \sum_{j=1}^{K_i} (N_i-1)^2 \left[\dfrac{\phi_i - \sigma_i^4}{N_i} + \dfrac{2\sigma_i^4}{N_i (N_i-1)} \right]}{\left[\sum_{i=1}^{c} K_i (N_i-1)\right]^2}$$

which at $N_i = \infty$ becomes

$$\text{Ult Var }(d) = \frac{\sum_{i=1}^{c} K_i \left[(N_i-1)\,(\phi_i - \sigma_i^4) + 2\sigma_i^4 \right]}{\left[\sum_{i=1}^{c} K_i\,(N_i-1)\right]\left[\sum_{i=1}^{c} K_i\,(N_i-1)\right]}$$

$$= 0$$

The validity of this result is perhaps somewhat easier to see when all $N_i = N$, in which case the (N-1)'s can be moved to the left of the summation signs and the single (N-1) in the numerator cancels with only one of the (N-1)'s in the denominator, leaving a still infinite denominator.

$$n = \frac{\sum_{i=1}^{c} N_i \sum_{j=1}^{K_i} (\overline{P}_{ij} - \overline{\overline{P}}_i)^2 + \sum_{i=1}^{c} N_i K_i (\overline{\overline{P}}_i - G)^2}{K-1}$$

$$\text{Var }(n) = \text{Var}\left[\frac{\sum_{i=1}^{c} N_i \sum_{j=1}^{K_i} (\overline{P}_{ij} - \overline{\overline{P}}_i)^2}{K-1}\right] + \text{Var}\left[\frac{\sum_{i=1}^{c} N_i K_i (\overline{\overline{P}}_i - G)^2}{K-1}\right]$$

$$\geq \text{Var}\left[\frac{\sum_{i=1}^{c} N_i \sum_{j=1}^{K_i} (\overline{P}_{ij} - \overline{\overline{P}}_i)^2}{K-1}\right]$$

$$\geq \text{Var}\left[\frac{\sum_{i=1}^{c} N_i (K_i-1)\, \hat{\sigma}^2_{\overline{P}_{ij}}}{K-1}\right]$$

$$\geq \sum_{i=1}^{c} \left[\frac{N_i (K_i-1)}{K-1}\right]^2 \text{Var }(\hat{\sigma}^2_{\overline{P}_{ij}})$$

Now let the smallest N_i approach infinity, (i.e., let all $N_i \to \infty$). Then, by the Central Limit Theorem, $\overline{P}_{ij}$ must be normally distributed if σ_i^2 is finite, thereby justifying the substitution of the normal-theory formula for Var ($\hat{\sigma}^2_{\overline{P}_{ij}}$).

$$\text{Ult Var (n)} \geq \sum_{i=1}^{c} \left[\frac{N_i\ (K_i-1)}{K-1} \right]^2 \left[\frac{2\sigma_{\bar{P}_{ij}}^4}{K_i-1} \right]$$

$$\geq \sum_{i=1}^{c} \left[\frac{N_i\ (K_i-1)}{K-1} \right]^2 \left[\frac{2\sigma_i^4}{N_i^2\ (K_i-1)} \right]$$

$$\geq \sum_{i=1}^{c} \frac{2\ (K_i-1)\ \sigma_i^4}{(K-1)^2} = \text{a nonzero constant, not involving } N_i.$$

Therefore

$$\frac{\text{Ult Var (d)}}{\text{Ult Var (n)}} = \frac{0}{\text{a constant}} = 0$$

Therefore, irrespective of normality or homogeneity, the ultimate denominator of F behaves as a constant relative to the ultimate numerator, and that constant is E (d). (See appendix VII.)

Now proceeding to show that the ultimate expected value of F has its normal theory value when variances are homogeneous or sample sizes are equal, but not generally otherwise:

$$F = \frac{\dfrac{\sum_{i=1}^{c} N_i \sum_{j=1}^{K_i} (\bar{P}_{ij}-\mu)^2 - T\ (\mu-G)^2}{K-1}}{\dfrac{\sum_{i=1}^{c} \sum_{j=1}^{K_i} (N_i-1)\ \hat{\sigma}_{ij}^2}{\sum_{i=1}^{c} K_i\ (N_i-1)}} = \frac{n}{d}$$

$$E\ (n) = \frac{\sum_{i=1}^{c} N_i \sum_{j=1}^{K_i} E\ (\bar{P}_{ij}-\mu)^2 - T\ E\ (\mu-G)^2}{K-1}$$

$$= \frac{\sum_{i=1}^{c} N_i \sum_{j=1}^{K_i} \sigma_{\bar{P}_{ij}}^2 - T\ \sigma_G^2}{K-1}$$

$$= \frac{\sum_{i=1}^{c} \sum_{j=1}^{K_i} \sigma_i^2 - T \sum_{i=1}^{c} K_i N_i \sigma_i^2 / T^2}{K-1}$$

$$= \frac{T \sum_{i=1}^{c} K_i \sigma_i^2 - \sum_{i=1}^{c} K_i N_i \sigma_i^2}{T (K-1)}$$

$$E (d) = \frac{\sum_{i=1}^{c} \sum_{j=1}^{K_i} (N_i - 1) \sigma_i^2}{\sum_{i=1}^{c} K_i (N_i - 1)} = \frac{\sum_{i=1}^{c} K_i (N_i - 1) \sigma_i^2}{\sum_{i=1}^{c} K_i (N_i - 1)}$$

If variances are homogeneous, so that all $\sigma_i^2 = \sigma^2$,

$$E (n) = \frac{\sigma^2 T \sum_{1}^{c} K_i - \sigma^2 \sum_{1}^{c} K_i N_i}{T (K-1)} = \frac{\sigma^2 T K - \sigma^2 T}{T (K-1)} = \sigma^2$$

$$E (d) = \frac{\sigma^2 \sum_{i=1}^{c} K_i (N_i - 1)}{\sum_{i=1}^{c} K_i (N_i - 1)} = \sigma^2$$

Thus E (n) = Normal-Theory E (n), E (d) = Normal-Theory E (d), and E (n) = E (d). Therefore

$$\text{Ultimate } E (F) = E \left[\frac{n}{E (d)} \right] = \frac{E (n)}{E (d)} = \frac{\sigma^2}{\sigma^2} = 1 = \text{Normal-Theory Ult } E (F).$$

If sample sizes are equal, so that all $N_i = N$, $T = \sum_1^c K_i N = N \sum_1^c K_i = N K$ **and**

$$E (n) = \frac{N K \sum_{i=1}^{c} K_i \sigma_i^2 - N \sum_{i=1}^{c} K_i \sigma_i^2}{N K (K-1)} = \frac{\sum_{i=1}^{c} K_i \sigma_i^2}{K}$$

$$E (d) = \frac{(N-1) \sum_{i=1}^{c} K_i \sigma_i^2}{(N-1) \sum_{i=1}^{c} K_i} = \frac{\sum_{i=1}^{c} K_i \sigma_i^2}{K} = E (n) \text{ and}$$

$$\text{Ult } E(F) = \frac{E(n)}{E(d)} = 1 = \text{Normal-Theory Ult } E(F).$$

At this point we have all the information necessary to determine the ultimate robustness of F under violation of the normality assumption alone. When all N_i become infinite, the denominator becomes functionally a constant, irrespective of normality or nonnormality, and that constant has the same value, σ^2, under normality as under nonnormality, so the denominator is entirely innocuous. The normality assumption actually required by the numerator is only that the sample means are normally distributed, since the numerator contains no original observations. But the fullfillment of this entirely equivalent substitute assumption is guaranteed by the Central Limit Theorem when all sample sizes are infinite, provided only that the common population variance is finite. Therefore when all N_i are infinite the normality assumption is met by the numerator and its failure in the denominator doesn't matter. (Not only does the numerator have its normal-theory mean and variance, but the shape of its distribution must be exactly its normal-theory shape, i.e., the Central Limit Theorem ensures that the numerator's distribution is identical to that expected under normal theory.) Consequently, under violation of the normality assumption alone, as N approaches infinity F approaches complete robustness, and any true significance level $\bar{p}$ approaches a limiting value of α.

We also have, at this point, enough information to state that the F test cannot be entirely ultimately robust under heterogeneity of variance when sample sizes differ, since under these conditions F does not even have its normal-theory mean.

The ultimate robustness of F under heterogeneity will be investigated only for the case of two possibly differing populations. This was the case investigated empirically in the present study, and derivations are somewhat simplified under this restriction. Actually, it can be proven that heterogeneity of variance prevents the F test from being ultimately completely robust regardless of the number of populations (with the single exception of the case where there are only two treatments and sample sizes are equal). However, it is intuitively reasonable that if heterogeneity of variance prevents complete ultimate robustness in the simpler two-population case, it should continue to do so when populations are added. Therefore generality will be sacrificed for conciseness and simplicity.

Let $P_1 = P$, $P_2 = Q$, $K_1 = K$, $K_2 = C$, $N_1 = N$, $N_2 = M$, ${\sigma_1}^2 = \sigma^2$, ${\sigma_2}^2 = R\,\sigma^2$, so that K samples, each of size N, are drawn from population P with mean μ and variance σ^2 and C samples, each of size M, are drawn from population Q with the same mean, μ, but with variance $R\,\sigma^2$ where R is the ratio of the Q population's variance to that of the P population. Then

$$G = \frac{\sum_1^c K_i N_i \bar{\bar{P}}_i}{\sum_1^c K_i N_i} = \frac{K N \bar{\bar{P}} + C M \bar{\bar{Q}}}{K N + C M}$$

$$\sum_1^c K_i N_i (\bar{\bar{P}}_i - G)^2 = K N \left(\bar{\bar{P}} - \frac{K N \bar{\bar{P}} + C M \bar{\bar{Q}}}{K N + C M}\right)^2 + C M \left(\bar{\bar{Q}} - \frac{K N \bar{\bar{P}} + C M \bar{\bar{Q}}}{K N + C M}\right)^2$$

$$= K N \left(\frac{C M \bar{\bar{P}} - C M \bar{\bar{Q}}}{K N + C M}\right)^2 + C M \left(\frac{K N \bar{\bar{Q}} - K N \bar{\bar{P}}}{K N + C M}\right)^2$$

$$= \frac{K N C M (C M + K N) (\bar{\bar{P}} - \bar{\bar{Q}})^2}{(K N + C M)^2} = \frac{K N C M (\bar{\bar{P}} - \bar{\bar{Q}})^2}{K N + C M}$$

$$F = \frac{\dfrac{\sum_{i=1}^c N_i \sum_{j=1}^{K_i} (\bar{P}_{ij} - \bar{\bar{P}}_i)^2 + \sum_{i=1}^c N_i K_i (\bar{\bar{P}}_i - G)^2}{\sum_{i=1}^c K_i - 1}}{\dfrac{\sum_{i=1}^c \sum_{j=1}^{K_i} \sum_{k=1}^{N_i} (P_{ijk} - \bar{P}_{ij})^2}{\sum_{i=1}^c K_i (N_i - 1)}}$$

$$= \frac{\dfrac{N \sum_1^K (\bar{P} - \bar{\bar{P}})^2 + M \sum_1^c (\bar{Q} - \bar{\bar{Q}})^2 + \frac{K N C M}{K N + C M} (\bar{\bar{P}} - \bar{\bar{Q}})^2}{K + C - 1}}{\dfrac{\sum_1^K \sum_1^N (P - \bar{P})^2 + \sum_1^C \sum_1^M (Q - \bar{Q})^2}{K (N-1) + C (M-1)}} = \frac{n}{d}$$

$$= \frac{\dfrac{N \sum_1^K (\bar{P}-\bar{\bar{P}})^2 + M \sum_1^c (\bar{Q}-\bar{\bar{Q}})^2 + \dfrac{K N C M}{K N + C M} (\bar{\bar{P}}-\bar{\bar{Q}})^2}{K + C - 1}}{\dfrac{\sum_1^K (N-1)\, \hat{\sigma}_P^{\,2} + \sum_1^c (M-1)\, \hat{\sigma}_Q^{\,2}}{K\,(N-1) + C\,(M-1)}}$$

Now let M and N both become infinite. In that case the denominator becomes its expected value, as already demonstrated, and since the mean of $\hat{\sigma}_X^{\,2}$ is $\sigma_X^{\,2}$ regardless of the distribution of X, no assumption of normality is introduced into the denominator. When N and M become infinite, $\bar{P}$, $\bar{\bar{P}}$, $\bar{Q}$, $\bar{\bar{Q}}$, and $\bar{\bar{P}}-\bar{\bar{Q}}$ all become normally distributed, due to the Central Limit Effect, the latter having a mean of $\mu - \mu = 0$ and a variance of $\sigma_{\bar{\bar{P}}-\bar{\bar{Q}}}^{\,2} = \sigma_{\bar{\bar{P}}}^{\,2} + \sigma_{\bar{\bar{Q}}}^{\,2}$. And, since the square of a normal variate with zero mean and unit variance is distributed as chi-square with one degree of freedom, $(\bar{\bar{P}}-\bar{\bar{Q}})^2/\sigma_{\bar{\bar{P}}-\bar{\bar{Q}}}^{\,2} = \chi_1^2$. Making this replacement and the other chi-square substitutions justified by the fact that $\sum_1^N (X-\bar{X})^2 = \chi_{N-1}^2\, \sigma_X^{\,2}$, and replacing the denominator by its expected value, the formula becomes, without making any assumption as to the normality of the original observations,

$$\text{Ult } F = \frac{\dfrac{N \chi_{K-1}^2 \sigma_{\bar{P}}^{\,2} + M \chi_{C-1}^2 \sigma_{\bar{Q}}^{\,2} + \dfrac{K N C M}{K N + C M} \chi_1^2 (\sigma_{\bar{\bar{P}}}^{\,2} + \sigma_{\bar{\bar{Q}}}^{\,2})}{K + C - 1}}{\dfrac{\sum_1^K (N-1)\, \sigma_P^{\,2} + \sum_1^c (M-1)\, \sigma_Q^{\,2}}{K\,(N-1) + C\,(M-1)}}$$

$$= \frac{\dfrac{\chi_{K-1}^2 \sigma_P^{\,2} + \chi_{C-1}^2 \sigma_Q^{\,2} + \chi_1^2 \left(\dfrac{C M \sigma_P^{\,2} + K N \sigma_Q^{\,2}}{K N + C M}\right)}{K + C - 1}}{\dfrac{K\,(N-1)\, \sigma_P^{\,2} + C\,(M-1)\, \sigma_Q^{\,2}}{K\,(N-1) + C\,(M-1)}}$$

$$= \frac{\dfrac{\sigma^2 \left[\chi_{K-1}^2 + \chi_{C-1}^2 R + \chi_1^2 \left(\dfrac{C M + K N R}{C M + K N}\right)\right]}{K + C - 1}}{\dfrac{\sigma^2 \,[K\,(N-1) + C\,R\,(M-1)]}{K\,(N-1) + C\,(M-1)}} = \frac{n}{d}$$

We already know that F cannot be completely ultimately robust when sample sizes are unequal because in that case ultimate F does not even have its normal-theory location. Therefore, let sample sizes be equal so that M = N. Then

$$\text{Ult } F = \frac{\dfrac{\sigma^2\left[\chi^2_{K-1} + \chi^2_{C-1} R + \chi^2_1 \left(\dfrac{C + K R}{C + K}\right)\right]}{K + C - 1}}{\dfrac{\sigma^2 (K + C R)}{K + C}} = \frac{n}{d}$$

$$= \frac{(K + C)\chi^2_{K-1} + R (K + C)\chi^2_{C-1} + (K R + C)\chi^2_1}{(K + C - 1)(K + C R)}$$

$$\text{Ult Var } (F) = \frac{(K + C)^2 (2 K - 2) + R^2 (K + C)^2 (2 C - 2) + (K R + C)^2 (2)}{(K + C - 1)^2 (K + C R)^2}$$

$$= \frac{2\left[(K + C)^2 (K + C R^2 - 1 - R^2) + K^2 R^2 + 2K C R + C^2\right]}{(K + C - 1)^2 (K + C R)^2}$$

$$= \frac{2\left[(K + C)^2 (K + C R^2) - (K^2 + C^2 R^2 + 2 K C R^2 + 2 K C) + 2 K C R\right]}{(K + C - 1)^2 (K + C R)^2}$$

$$= \frac{2\left[(K + C)(K^2 + C^2R^2 + KCR^2 + KC) - (K^2 + C^2R^2 + 2KCR^2 + 2KC - 2KCR)\right]}{(K + C - 1)^2 (K + C R)^2}$$

$$= \frac{2\left[(K + C - 1)(K^2 + C^2 R^2 + K C R^2 + K C) - K C (R^2 - 2R + 1)\right]}{(K + C - 1)^2 (K + C R)^2}$$

$$= \frac{2\left[(K + C - 1)(K^2 + C^2 R^2) + K C (K + C - 1)(R^2 + 1) - K C (R - 1)^2\right]}{(K + C - 1)^2 (K + C R)^2}$$

$$= \frac{2\left[(K + C - 1)(K + C R)^2 + K C (K + C - 2)(R - 1)^2\right]}{(K + C - 1)^2 (K + C R)^2}$$

The ultimate normal-theory variance of F is 2 divided by the number of degrees of freedom in its numerator. This same value of $2/(K + C - 1)$ is obtained by setting $R = 1$ in the above formula. Therefore

$$\frac{\text{Ult Var (F)}}{\text{Ult Norm Theory Var (F)}} = \frac{\dfrac{2\left[(K + C - 1)(K + C\,R)^2 + K\,C\,(K + C - 2)(R - 1)^2\right]}{(K + C - 1)^2\,(K + C\,R)^2}}{\dfrac{2}{K + C - 1}}$$

$$= \frac{(K + C - 1)(K + C\,R)^2 + K\,C\,(K + C - 2)(R - 1)^2}{(K + C - 1)\,(K + C\,R)^2}$$

$$= 1 + K\,C\left(\frac{K + C - 2}{K + C - 1}\right)\left(\frac{R - 1}{K + C\,R}\right)^2$$

Since the E (F) = Normal Theory E (F), in this case where sample sizes are equal the above ratio of ultimate variances serves as a direct index of ultimate robustness. Notice that since neither K nor C can be less than 1 and since a squared term must be positive, the ultimate ratio can exceed 1 but cannot be less than 1. Furthermore, only when $R = 1$ or when $K = C = 1$ does the ratio equal 1. Therefore, under heterogeneity of variance with equal sample sizes the ultimate variance of multi-sample F is always larger than that under normal theory, and since the expected value of F is the same as under normal theory, one would expect $\bar{p}$ to exceed α at the testing tails, as indeed it did in the present study.

If we let $C = \gamma K$, so that γ is the ratio of the number of samples drawn from the Q population to the number drawn from the P population (just as R is σ_Q^2/σ_P^2), the ultimate ratio becomes

$$\frac{\text{Ult Var (F)}}{\text{Ult Norm Theory Var (F)}} = 1 + \gamma\,K^2\left(\frac{K + \gamma\,K - 2}{K + \gamma\,K - 1}\right)\left(\frac{R - 1}{K + \gamma\,K\,R}\right)^2$$

$$= 1 + \gamma\left(\frac{\gamma\,K + K - 2}{\gamma\,K + K - 1}\right)\left(\frac{R - 1}{\gamma\,R + 1}\right)^2$$

And, taking the square roots of both sides of the equation, we obtain an index of ultimate robustness in terms of (perhaps more easily visualized and interpreted) standard deviations of the actual vs. normal-theory F distributions.

$$\frac{\text{Ult Stand. Dev. (F)}}{\text{Ult. Norm Theory Stand. Dev. (F)}} = \sqrt{1 + \gamma\left(\frac{\gamma K + K - 2}{\gamma K + K - 1}\right)\left(\frac{R - 1}{\gamma R \quad 1}\right)^2}$$

This ratio is plotted as ordinate against $\sqrt{R}$, i.e., the ratio between the standard deviations of the two sampled populations, as abscissa in Figures 55-60 for various constant values of γ and K. A connection between theoretical and empirical results is afforded by the fact that, for infinite N, FYB N,NN and FYB N,NNN correspond, in Figures 58-60, to an abscissa of $\sqrt{R} = 1/2$ on the curves labeled $\gamma = 2$ and $\gamma = 3$, respectively, and FYB NN,N and FYB NNN,N are represented by points on the same respective curves whose abscissa is $\sqrt{R} = 2$. A rough idea of the relationship between the ratio $\frac{\text{Ult Stand. Dev. (F)}}{\text{Ult Norm Theory Stand. Dev. (F)}}$ and the ratio α/ρ can be gotten by comparing ordinates of the four above-mentioned points in Figures 58-60 with ordinates on the graphs giving the empirical α/ρ ratios for the four FYB statistics at N = 1024.

Figures 55 to 60 only show curves for $\gamma \geq 1$, and this is all that is necessary since by letting R be the relative variance of the population from which the larger number of samples is drawn γ can be forced into the range of values ≥ 1. When $\gamma = 1$, i.e., when equal numbers of samples are drawn from the two populations, it is entirely arbitrary whether $R = \sigma_1^2/\sigma_2^2$ or σ_2^2/σ_1^2, i.e., whether σ_1^2/σ_2^2 equals R or 1/R. The same ratio of ultimate F standard deviations is obtained in either case. Consequently the curves for $\gamma = 1$, $R = \Delta$ (where $1 \leq \Delta \leq \infty$) should have essentially the same appearance as, i.e., should be the mirror images of, the curves for $\gamma = 1$, $R = 1/\Delta$. This requirement was met by using mutually reciprocal abscissa scales on opposite sides of 1. The net result is that a reciprocal abscissa scale is used on one side of 1 and the reciprocal of a reciprocal scale, i.e., a linear abscissa scale, is used on the other.

Figures 55-60 show some interesting effects, which can also be deduced from the formula upon which they are based. If γ and R are held constant, the ratio of ultimate F standard deviations (which, for brevity, we will call the nonrobustness

SENSITIVITY OF F TO HETEROGENEITY OF VARIANCE WHEN SAMPLE SIZES ARE EQUAL AND INFINITE

RATIO OF ULTIMATE ($N=\infty$) STANDARD DEVIATION OF F DISTRIBUTION TO ULTIMATE NORMAL-THEORY ($R=1$) STANDARD DEVIATION OF F, WHEN F STATISTIC IS BASED UPON k SAMPLES, EACH OF SIZE N, FROM A NORMAL POPULATION WITH VARIANCE σ^2 AND γk SAMPLES, EACH OF SIZE N, FROM A NORMAL POPULATION WITH THE SAME MEAN BUT WITH VARIANCE $R\sigma^2$

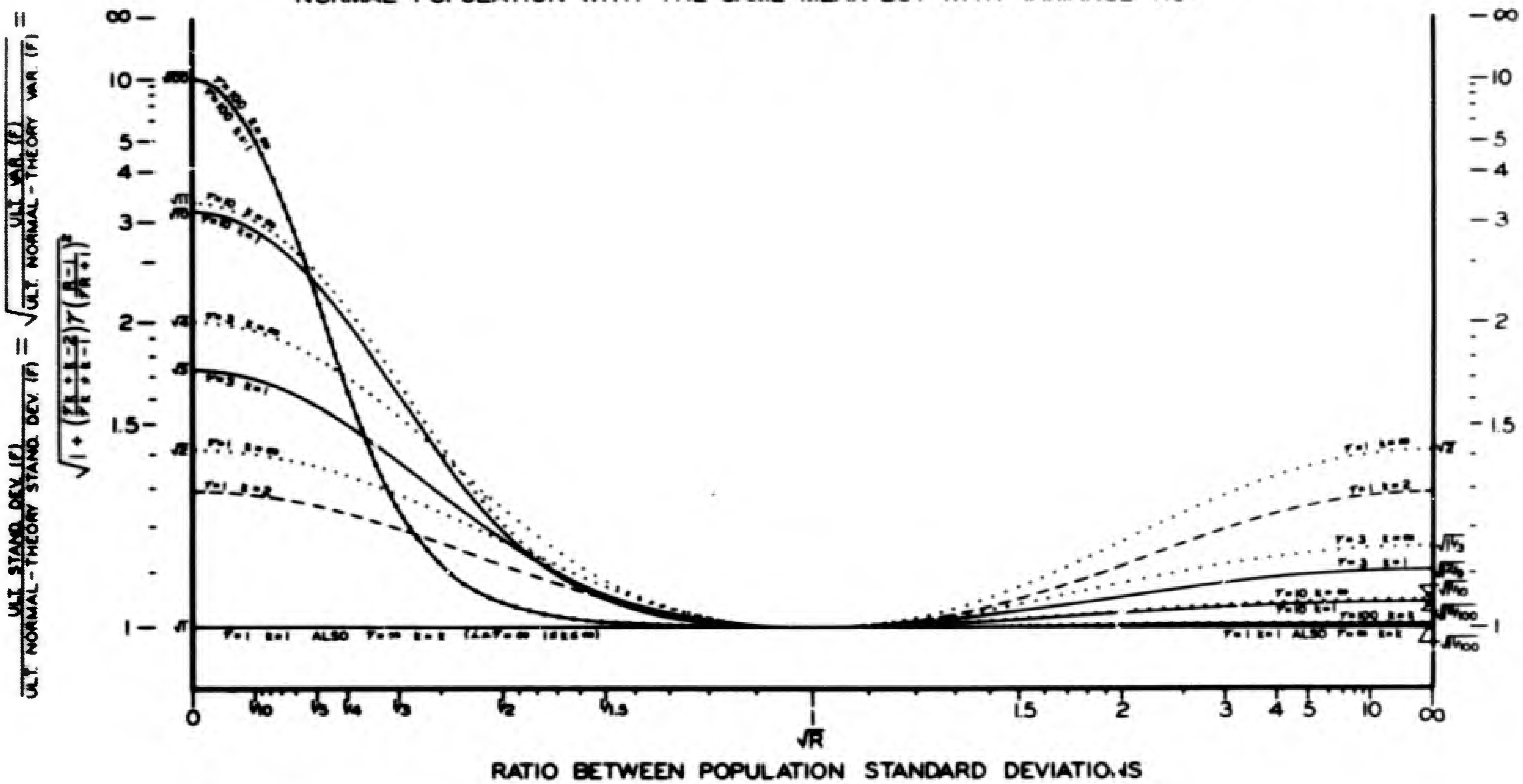

Figure 55. Ultimate Robustness of F when Samples are of Equal Size and There are No More than Two Different Population Variances

SENSITIVITY OF F TO HETEROGENEITY OF VARIANCE WHEN SAMPLE SIZES ARE EQUAL AND INFINITE

RATIO OF ULTIMATE ($N=\infty$) STANDARD DEVIATION OF F DISTRIBUTION TO ULTIMATE NORMAL-THEORY ($R=1$) STANDARD DEVIATION OF F, WHEN F STATISTIC IS BASED UPON K SAMPLES, EACH OF SIZE N, FROM A NORMAL POPULATION WITH VARIANCE σ^2 AND γK SAMPLES, EACH OF SIZE N, FROM A NORMAL POPULATION WITH THE SAME MEAN BUT WITH VARIANCE $R\sigma^2$

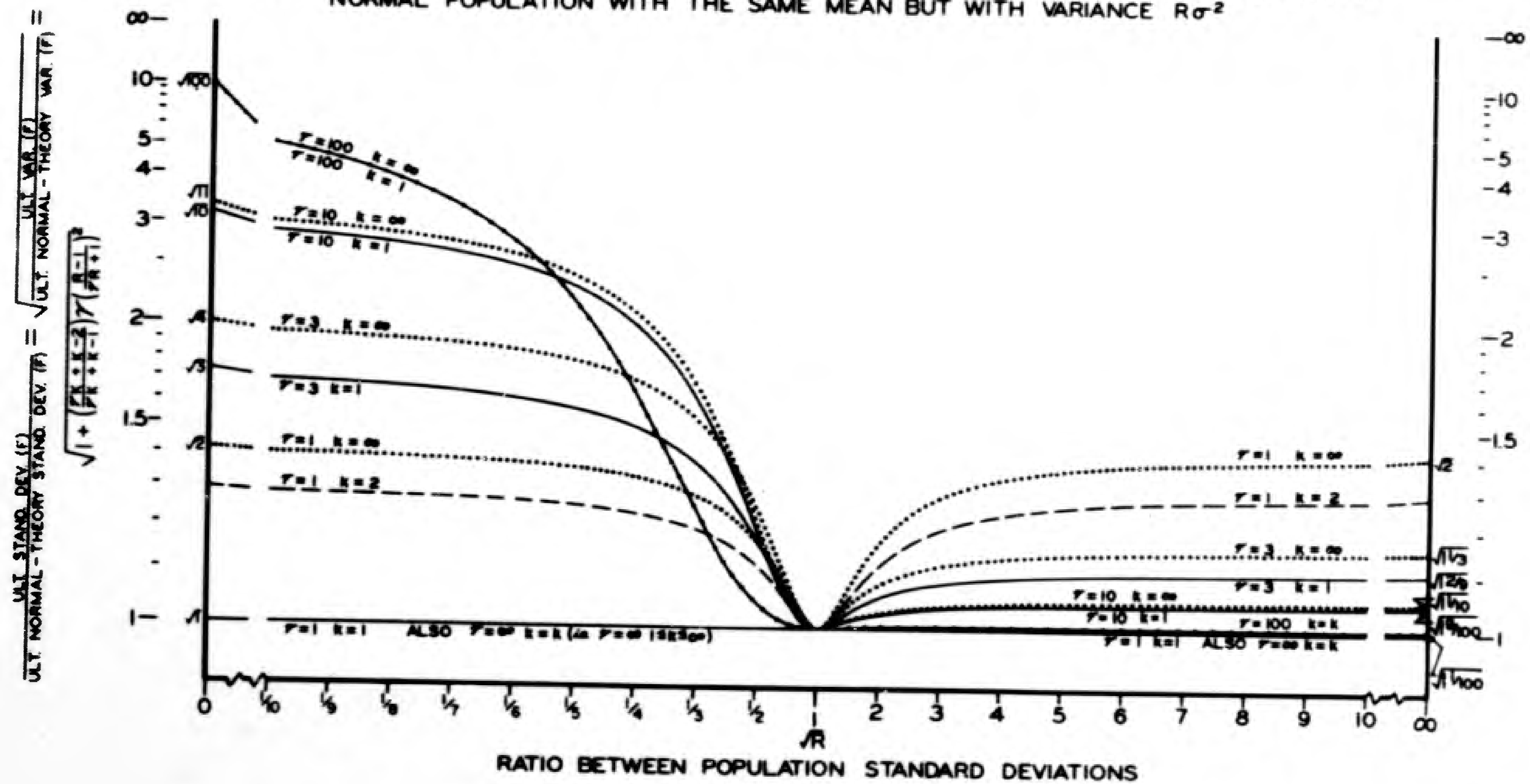

Figure 56. Same Data as Figure 55, Plotted Different Abscissa Scale

SENSITIVITY OF F TO HETEROGENEITY OF VARIANCE WHEN SAMPLE SIZES ARE EQUAL AND INFINITE

RATIO OF ULTIMATE ($N=\infty$) STANDARD DEVIATION OF F DISTRIBUTION TO ULTIMATE NORMAL-THEORY ($R=1$) STANDARD DEVIATION OF F, WHEN F STATISTIC IS BASED UPON k SAMPLES, EACH OF SIZE N, FROM A NORMAL POPULATION WITH VARIANCE σ^2 AND γk SAMPLES, EACH OF SIZE N, FROM A NORMAL POPULATION WITH THE SAME MEAN BUT WITH VARIANCE $R\sigma^2$

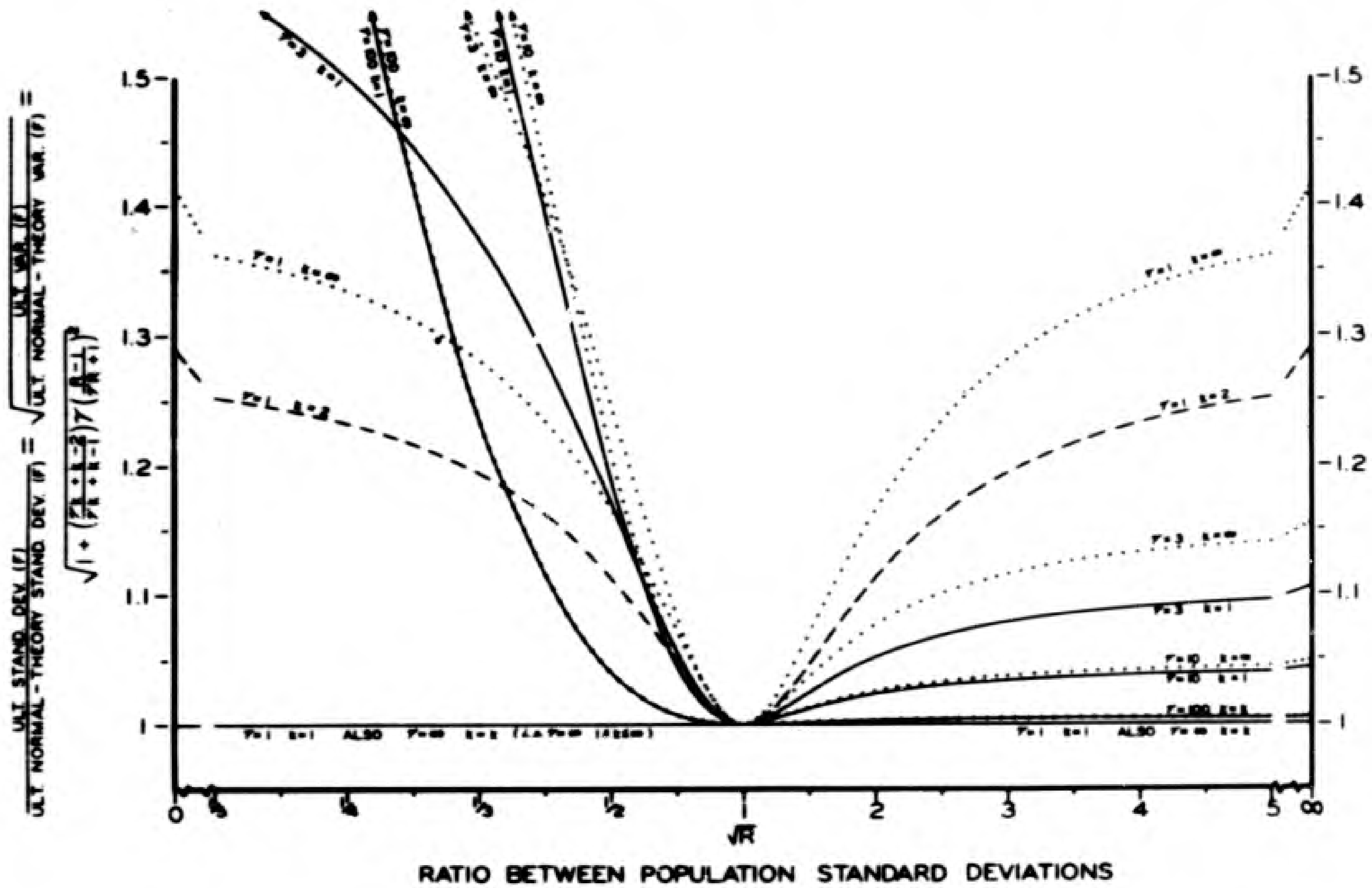

Figure 57. Critical Section of Figure 56, Plotted to Linear Ordinate Scale

index) increases with increasing values of K and the curves for $K = 1$ and $K = \infty$ form boundaries for the region within which all of the other K curves having the same γ will lie. The influence of K diminishes as γ exceeds 1 by greater and greater amounts, and the region enclosed between the $K = 1$ and $K = \infty$ curves for the same γ becomes increasingly narrow. Also, for fixed values of K and γ (other than $\gamma = 1$, $K = 1$ or $\gamma = \infty$, $K = K$) the nonrobustness index increases as R increases above 1 or decreases below 1. For $K = 1$ it reaches an absolute maximum of $\sqrt{\gamma}$ at $R = 0$ and a local maximum for values of R greater than 1 of $\sqrt{1 + \frac{\gamma - 1}{\gamma^2}}$ at R = infinity, assuming in both cases that $\gamma > 1$. For $K = \infty$ an entirely analogous statement applies with respective maxima of $\sqrt{1 + \gamma}$ and $\sqrt{1 + \frac{1}{\gamma}}$. However, while the nonrobustness index increases monotonically with increasing values of K or with unidirectionally increasing departures of R from 1, its behavior with regard to γ is far more complex. The curves in Figures 55-60 suggest that if K and R are held constant while γ is allowed to increase (above $\gamma = 2$), the nonrobustness index decreases monotonically when R is greater than 1 (at least for the larger values of R), but increases until γ reaches some critical value and then decreases when R is less than 1. Furthermore, the critical value of γ appears to depend upon R.

Explicit information on these matters can be obtained by taking the derivative of the robustness index with respect to γ. For values of γ which make the derivative positive, increases in γ cause the robustness index to increase; and for values of γ which make the derivative negative, increases in γ cause the robustness index to decrease. Therefore if the derivative is zero for a single value of γ, has one algebraic sign for all larger values of γ and the opposite sign for all smaller values, that single value of γ for which the derivative is zero is the critical value and as γ departs increasingly from it in a single direction the robustness index decreases monotonically. Instead of differentiating the entire robustness index, the same end is accomplished more simply by differentiating only the variable part of the expression under the radical, $(R-1)^2$ being now regarded as a constant:

SENSITIVITY OF F TO HETEROGENEITY OF VARIANCE WHEN SAMPLE SIZES ARE EQUAL AND INFINITE

RATIO OF ULTIMATE ($N=\infty$) STANDARD DEVIATION OF F DISTRIBUTION TO ULTIMATE NORMAL-THEORY ($R=1$) STANDARD DEVIATION OF F, WHEN F STATISTIC IS BASED UPON ONE SAMPLE, OF SIZE N, FROM A NORMAL POPULATION WITH VARIANCE σ^2 AND τ SAMPLES, EACH OF SIZE N, FROM A NORMAL POPULATION WITH THE SAME MEAN BUT WITH VARIANCE $R\sigma^2$

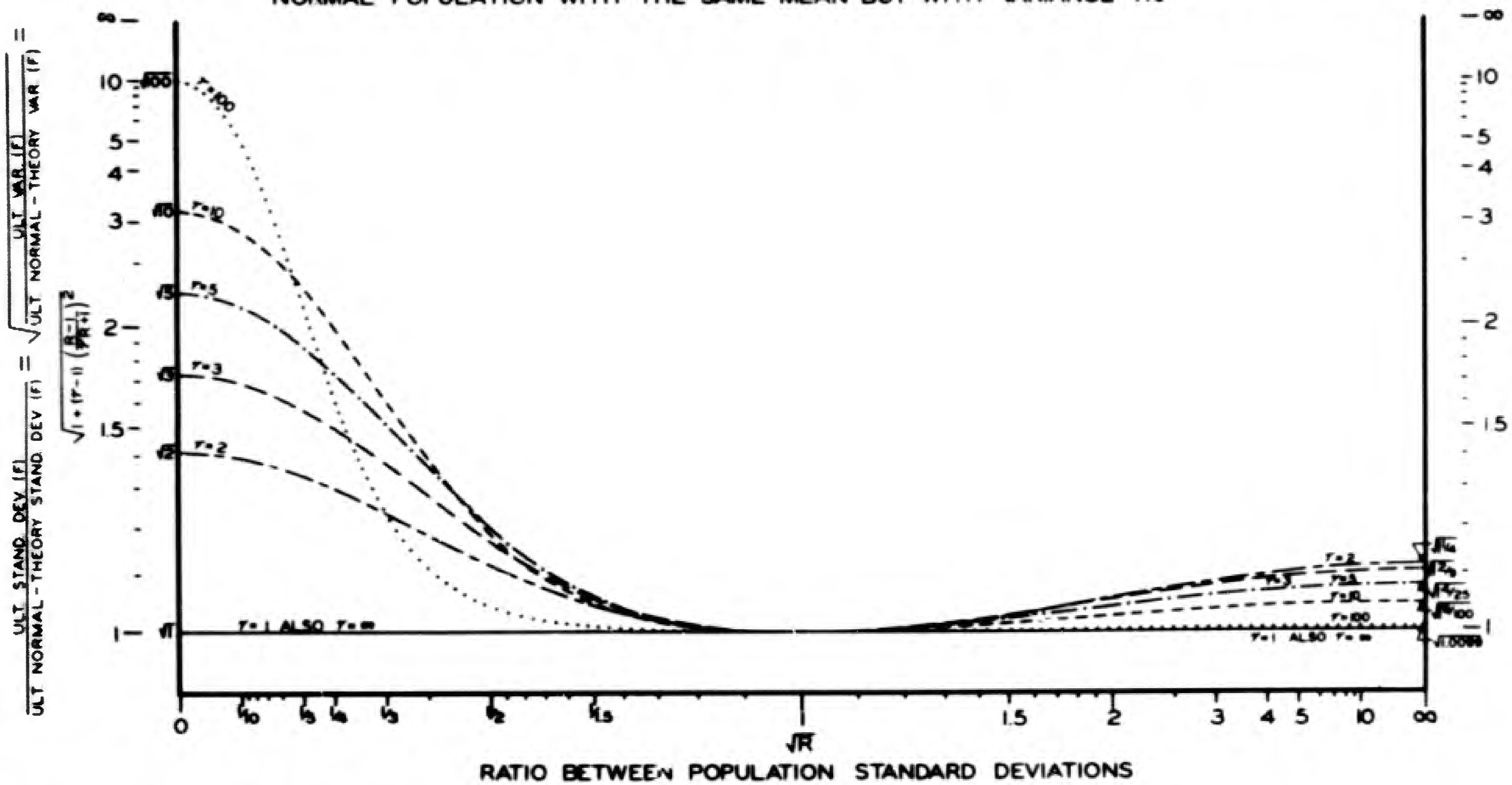

Figure 58. Ultimate Robustness of F when Samples are of Equal Size and all Samples but One Come from Populations with the Same Variance

SENSITIVITY OF F TO HETEROGENEITY OF VARIANCE WHEN SAMPLE SIZES ARE EQUAL AND INFINITE

RATIO OF ULTIMATE (N=∞) STANDARD DEVIATION OF F DISTRIBUTION TO ULTIMATE NORMAL-THEORY (R=1) STANDARD DEVIATION OF F, WHEN F STATISTIC IS BASED UPON ONE SAMPLE, OF SIZE N, FROM A NORMAL POPULATION WITH VARIANCE σ^2 AND r SAMPLES, EACH OF SIZE N, FROM A NORMAL POPULATION WITH THE SAME MEAN BUT WITH VARIANCE $R\sigma^2$

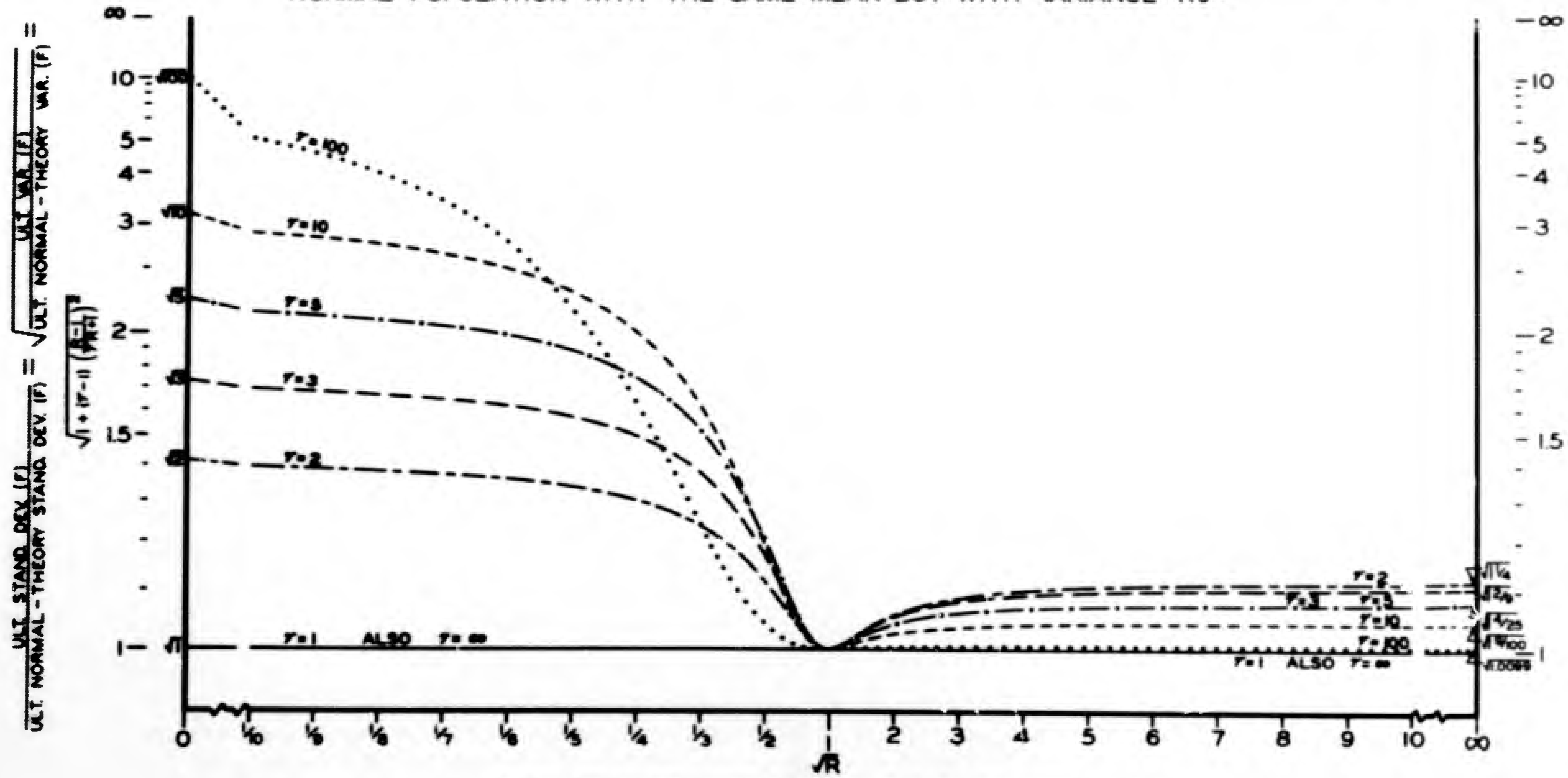

Figure 59. Same Data as Figure 58, Plotted to Different Abscissa Scale

SENSITIVITY OF F TO HETEROGENEITY OF VARIANCE WHEN SAMPLE SIZES ARE EQUAL AND INFINITE

RATIO OF ULTIMATE ($N=\infty$) STANDARD DEVIATION OF F DISTRIBUTION TO ULTIMATE NORMAL-THEORY ($R=1$) STANDARD DEVIATION OF F, WHEN F STATISTIC IS BASED UPON ONE SAMPLE, OF SIZE N, FROM A NORMAL POPULATION WITH VARIANCE σ^2 AND r SAMPLES, EACH OF SIZE N, FROM A NORMAL POPULATION WITH THE SAME MEAN BUT WITH VARIANCE $R\sigma^2$

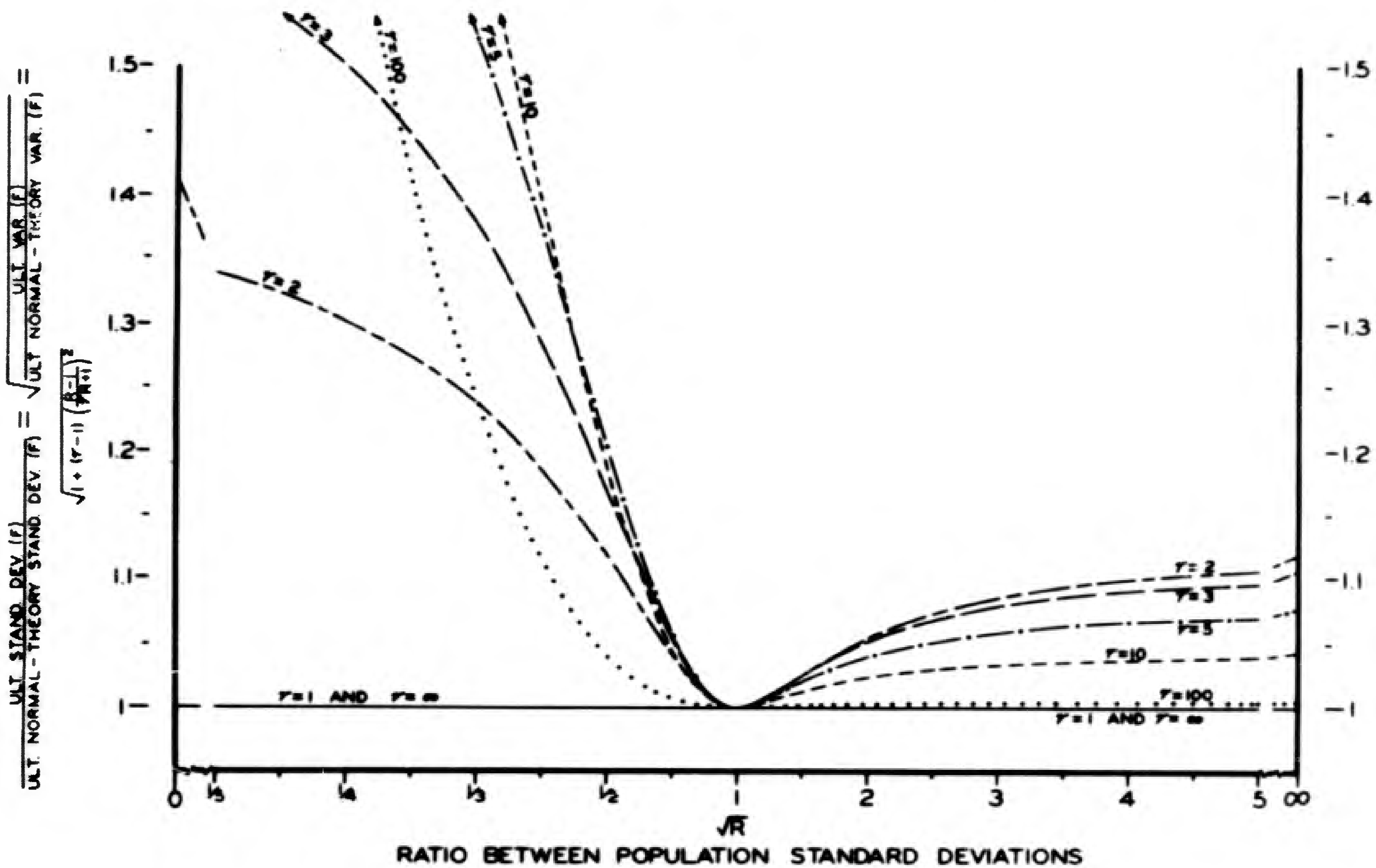

Figure 60. Critical Section of Figure 59, Plotted to Linear Ordinate Scale

$$\frac{d}{d\gamma}\left(\frac{\gamma K + K - 2}{\gamma K + K - 1}\right)\left(\frac{\gamma}{(\gamma R + 1)^2}\right) = \left(\frac{\gamma K + K - 2}{\gamma K + K - 1}\right)\left(\frac{(\gamma R + 1)^2 - 2\gamma R(\gamma R + 1)}{(\gamma R + 1)^4}\right)$$

$$+ \left(\frac{\gamma}{(\gamma R + 1)^2}\right)\left(\frac{(\gamma K + K - 1) K - (\gamma K + K - 2) K}{(\gamma K + K - 1)^2}\right)$$

$$= \left(\frac{\gamma K + K - 2}{\gamma K + K - 1}\right)\left(\frac{1 - \gamma R}{(\gamma R + 1)^3}\right) + \frac{\gamma K}{(\gamma R + 1)^2(\gamma K + K - 1)^2}$$

$$= \frac{(\gamma K + K - 1)(\gamma K + K - 2)(1 - \gamma R) + \gamma K(\gamma R + 1)}{(\gamma K + K - 1)^2(\gamma R + 1)^3}$$

If we let $K = \infty$, the derivative reduces to $\frac{1 - \gamma R}{(\gamma R + 1)^3}$. Since γ and R are always positive, the algebraic sign of the derivative is determined by its numerator, which is positive when $1 > \gamma R$, i.e., when $\gamma < 1/R$, and negative when the inequality sign is reversed. The entire derivative is zero when $\gamma R = 1$, i.e., when $\gamma = 1/R$. Letting $K = 1$ reduces the derivative to

$$\frac{\gamma(\gamma-1)(1-\gamma R) + \gamma(\gamma R+1)}{\gamma^2(\gamma R+1)^3} = \frac{\gamma - \gamma^2 R - 1 + \gamma R + \gamma R + 1}{\gamma(\gamma R+1)^3} = \frac{1 - \gamma R + 2R}{(\gamma R+1)^3}.$$

Again the algebraic sign is determined by the numerator. It is positive when $1 + 2R > \gamma R$, i.e., when $\gamma < (2R + 1)/R$ and negative when the inequality sign is reversed. The entire derivative is zero when $\gamma R = 1 + 2R$, i.e., when $\gamma = (2R + 1)/R$. Thus, if γ is increased from 0 to ∞ while R and K are held constant, the nonrobustness index increases monotonically until γ reaches its critical value after which the nonrobustness index decreases monotonically toward a value of 1 (indicating complete robustness), which it assumes when $\gamma = \infty$. The critical value is $\gamma = (2R + 1)/R$ when $K = 1$ and is $\gamma = 1/R$ when $K = \infty$. (It should be borne in mind, however, that the part of the monotonic increase in the nonrobustness index corresponding to $\gamma < \frac{1}{K}$ is meaningless since γ K, the number of samples drawn from the population with variance R σ^2, must be an integer ≥ 1.)

We previously found that for $\gamma \geq 1$ the robustness index reaches an absolute maximum at R = 0 of $\sqrt{\gamma}$ when K = 1 and $\sqrt{1 + \gamma}$ when K = ∞ and for R values greater than 1 reaches a maximum at R = ∞ of $\sqrt{1 + \frac{\gamma - 1}{\gamma^2}}$ when K = 1 and $\sqrt{1 + \frac{1}{\gamma}}$ when K = ∞. These maxima are all in terms of γ at known values of R and K. But we now have the maximum value of γ in terms of R for fixed values of R and K. Therefore we may obtain the actual numerical values of the former maxima at the critical values of γ, and these will be the largest values the robustness index can assume for any γ. Letting I stand for the robustness index, γ' stand for the critical value of γ, and I' stand for the robustness index when γ assumes its critical value, we have for $\gamma \geq 1$:

At R = 0, K = 1

$$\gamma' = \frac{2R + 1}{R} = \frac{0 + 1}{0} = \infty$$

$$I' = \sqrt{\gamma'} = \infty$$

At R = 0, K = ∞

$$\gamma' = \frac{1}{R} = \frac{1}{0} = \infty$$

$$I' = \sqrt{1 + \gamma'} = \sqrt{1 + \infty} = \infty$$

At R = ∞, K = 1

$$\gamma' = \frac{2R + 1}{R} = \frac{2\infty + 1}{\infty} = 2$$

$$I' = \sqrt{1 + \frac{\gamma' - 1}{\gamma'^2}} = \sqrt{1 + \frac{1}{4}} = 1.118$$

At R = ∞, K = ∞

$$\gamma' = \frac{1}{R} = \frac{1}{\infty} = 0$$

$$I' = \sqrt{1 + \frac{1}{\gamma'}} = \sqrt{1 + \frac{1}{0}} = \sqrt{1 + \infty} = \infty$$

Thus we are faced with the awesome realization that even if we equate sample sizes and base each sample upon an infinite number of observations, thereby ensuring that the F distribution will have its normal-theory mean, there is still no absolute, finite, unqualified upper bound to the possible nonrobustness of the F test. To be sure the conditions, e.g., $R = 0$, $R = \infty$, upon which the attainment of infinite nonrobustness depend are unrealistic, but then so are the protective conditions, i.e., infinite and equal sample sizes, under which we have investigated F in obtaining the damning evidence. One would hope that when investigated under gratuitously advantageous conditions, a test widely reputed to be robust would at least be found to have a finite, unqualified, upper bound to its nonrobustness. Some comfort may be derived from the fact that I has a very low upper bound of 1.118 when $K = 1$ provided that $1 \leq R \leq \infty$; however the usefulness of this knowledge is limited by the facts that the bound does not apply when $0 \leq R < 1$ and that the experimenter can seldom be expected to know whether R is greater or less than 1.

Now let us return to the formula

$$F = \frac{\dfrac{N \sum_1^K (\bar{P}-\bar{\bar{P}})^2 + M \sum_1^C (\bar{Q}-\bar{\bar{Q}})^2 + \frac{K\,N\,C\,M}{K\,N + C\,M} (\bar{\bar{P}}-\bar{\bar{Q}})^2}{K + C - 1}}{\dfrac{\sum_1^K \sum_1^N (P-\bar{P})^2 + \sum_1^C \sum_1^M (Q-\bar{Q})^2}{K\,(N-1) + C\,(M-1)}}$$

and let both K and C equal one. In that case, $\bar{\bar{P}} = \sum_1^K \bar{P}/K = \sum_1^1 \bar{P}/1 = \bar{P}$

and, likewise, $\bar{\bar{Q}} = \bar{Q}$, so that $\bar{P}-\bar{\bar{P}} = 0$ and $\bar{Q}-\bar{\bar{Q}} = 0$, the formula becoming

$$F = \frac{\frac{N\,M}{N + M} (\bar{P}-\bar{Q})^2}{\dfrac{\sum_1^N (P-\bar{P})^2 + \sum_1^M (Q-\bar{Q})^2}{N-1 + M-1}}$$

$$= \frac{(\bar{P}-\bar{Q})^2}{\dfrac{\sum_1^N (P-\bar{P})^2 + \sum_1^M (Q-\bar{Q})^2}{N + M-2} \left(\dfrac{N + M}{N\,M}\right)}$$

$$= \frac{(\bar{P}-\bar{Q})^2}{\frac{\sum_1^N (P-\bar{P})^2 + \sum_1^M (Q-\bar{Q})^2}{N + M-2} \left(\frac{1}{N} + \frac{1}{M}\right)}$$

$$= t^2$$

So, treating F and t as discrete variables in order to simplify notation,

$\sum_{F_\alpha}^{+\infty} \Pr(F) = \sum_{t_\alpha^2}^{+\infty} \Pr(t^2)$. But the probability of t^2 is the same as the probability of $|t|$, so

$\sum_{F_\alpha}^{+\infty} \Pr(F) = \sum_{|t|_\alpha}^{+\infty} \Pr(|t|)$. In other words, when based upon two samples, the F test using an upper-tail rejection region corresponding to a normal-theory significance level of α (i.e., applied in conventional fashion) is equivalent to a two-sample t test using a symmetrical two-tailed rejection region corresponding also to a normal-theory significance level of α. Therefore the robustness of F when K = C = 1 is equally the robustness of the two-tailed t test and the graphs and data already presented for the two-tailed, two-sample t test may, with equal validity be regarded as graphs and data for the two-sample F test.

Since the case of K = C = 1 is an important special case and one for which we have a great deal of empirical data, it is worth while to investigate the ultimate robustness of F under heterogeneity in this case when sample sizes are unequal (even though we know in advance that F does not have its normal-theory location and therefore cannot be ultimately completely robust). Letting K = C = 1 we have

$$F = \frac{\frac{NM}{N + M} (\bar{P}-\bar{Q})^2}{\frac{\sum_1^N (P-\bar{P})^2 + \sum_1^M (Q-\bar{Q})^2}{N-1 + M-1}} = \frac{\frac{NM}{N + M} (\bar{P}-\bar{Q})^2}{\frac{(N-1)\,\hat{\sigma}_P^2 + (M-1)\,\hat{\sigma}_Q^2}{N + M-2}}$$

Now letting both N and M approach infinity, we can replace the denominator by its expected value, and the Central Limit Theorem ensures the normality of $\bar{P}-\bar{Q}$ with variance $\sigma_{\bar{P}-\bar{Q}}^2 = \sigma_{\bar{P}}^2 + \sigma_{\bar{Q}}^2$ so that

$$\text{Ult } F = \frac{\frac{NM}{N+M}\,\chi_1^2\,(\sigma_{\bar{P}}^2 + \sigma_{\bar{Q}}^2)}{\frac{(N-1)\,\sigma_P^2 + (M-1)\,\sigma_Q^2}{N+M-2}}$$

$$= \frac{\chi_1^2\left(\frac{M\sigma^2 + NR\sigma^2}{N+M}\right)}{\frac{(N-1)\,\sigma^2 + (M-1)\,R\sigma^2}{N+M-2}}$$

which, if we let $M = rN$, becomes

$$= \frac{\chi_1^2\,\sigma^2\left(\frac{r+R}{1+r}\right)}{\frac{\sigma^2(N+rNR-1-R)}{N+rN-2}}$$

and, since N is infinite, the denominator is

$$d = \frac{\sigma^2 N\,(1+rR)}{N\,(1+r)-2} - \frac{\sigma^2\,(R+1)}{N\,(1+r)-2}$$

$$= \frac{\sigma^2\,(1+rR)}{1+r-(2/N)} - 0 = \sigma^2\left(\frac{1+rR}{1+r}\right) \quad \text{so that}$$

$$\text{Ult } F = \frac{\chi_1^2\,\sigma^2\left(\frac{r+R}{r+1}\right)}{\sigma^2\left(\frac{rR+1}{r+1}\right)} = \chi_1^2\left(\frac{r+R}{rR+1}\right)$$

$$\text{Ult } E(F) = \left(\frac{r+R}{rR+1}\right)$$

whereas the normal-theory Ult E(F), obtained by setting $R = 1$, is 1, and

$$\text{Ult Var }(F) = 2\left(\frac{r+R}{rR+1}\right)^2$$

whereas the normal-theory ultimate Var (F) is 2. Thus $(r + R)/(rR + 1)$ is both the ratio of the ultimate expected value of F to the ultimate normal-theory expected value and the ratio of the ultimate standard deviation of F to the ultimate normal-theory standard deviation. Both ratios tell something important about the ultimate robustness of F under heterogeneity, but neither tells the entire story. When sample sizes were equal, so were ultimate expected values under normal-theory and otherwise, and therefore the ratio of ultimate variances under the two situations adequately described the effect of heterogeneity (with or without nonnormality) upon the ultimate robustness of F. However, since under heterogeneity when samples are of unequal size Ult E (F) $\neq$ Ult Normal-Theory E (F), any index of ultimate robustness to be self-sufficient will have to take account of the bias in Ult E (F), due to heterogeneity, as well as of the resulting bias in Ult Var (F). An intuitively meaningful index which does so is the ratio of the ultimate expected value of F plus Δ ultimate standard deviations of F to the ultimate normal-theory expected value of F plus Δ ultimate normal-theory standard deviations of F (where Δ can be any number and can be negative as well as positive). Letting SD (F) stand for standard deviation of F, the index is

$$\frac{\text{Ult E (F)} + \Delta[\text{Ult SD (F)}]}{\text{Ult Normal-Theory E (F)} + \Delta[\text{Ult Normal-Theory SD (F)}]} = \frac{\left(\frac{r + R}{rR + 1}\right) + \Delta\sqrt{2}\left(\frac{r + R}{rR + 1}\right)}{1 + \Delta\sqrt{2}}$$

$$= \frac{(1 + \Delta\sqrt{2})\left(\frac{r + R}{rR + 1}\right)}{1 + \Delta\sqrt{2}}$$

$$= \frac{r + R}{rR + 1}$$

which is independent of the value of Δ and which is the same value obtained for the ratio between ultimate means and for the ratio between ultimate standard deviations.

However, the ratio $(r + R)/(rR + 1)$ is an even more concisely meaningful index of the ultimate robustness of t than of $t^2 = F$. Consider, again, the statistic

$$F = t^2 = \frac{\frac{NM}{N + M}(\bar{P}-\bar{Q})^2}{\frac{(N-1)\hat{\sigma}_P^2 + (M-1)\hat{\sigma}_Q^2}{N + M-2}} = \frac{\frac{rN}{1 + r}(\bar{P}-\bar{Q})^2}{\frac{(N-1)\hat{\sigma}_P^2 + (rN-1)\hat{\sigma}_Q^2}{N + rN-2}} = \frac{n}{d}$$

in the second form of which M has been replaced by rN. We already know that as N approaches infinity, Var (d)/Var (n) approaches zero so that, relative to n, d is a constant, E (d). So, ultimately, t^2 is simply the ratio of n to a constant, E (d), and

$$\text{Ult } t = \sqrt{\frac{n}{d}} = \sqrt{\frac{n}{E\ (d)}} = \sqrt{\frac{\frac{rN}{1 + r} \quad (\bar{P}-\bar{Q})^2}{\frac{(N-1)\ \sigma^2 + (rN-1)\ R\ \sigma^2}{N + rN-2}}}$$

$$= \frac{\bar{P}-\bar{Q}}{\sqrt{\frac{\sigma^2\ (1 + r)(N-1 + rNR - R)}{rN(N + rN-2)}}} = \frac{n'}{d'}$$

where d' is a constant. Therefore,

$$\text{Var (t)} = \frac{\text{Var (n')}}{(d')^2} = \frac{\sigma^2_{\bar{P}-\bar{Q}}}{\frac{\sigma^2(1 + r)(N-1 + rNR - R)}{rN\ (N + rN-2)}}$$

$$= \frac{\frac{\sigma^2}{N} + \frac{R\sigma^2}{rN}}{\frac{\sigma^2\ (1 + r)(N + rNR - 1 - R)}{rN\ (N + rN-2)}} = \frac{\frac{r + R}{rN}}{\frac{(1 + r)(N + rNR - 1 - R)}{rN\ (N + rN-2)}}$$

$$\text{and Ult Var (t)} = \frac{r + R}{(1 + r)\left(\frac{1 + rR}{1 + r - (2/N)} - \frac{R + 1}{N + rN-2}\right)} = \frac{r + R}{1 + rR}$$

and the ultimate normal-theory Var (t), obtained by setting R = 1 in the above formula, is 1, so

$$\frac{\text{Ult Var (t)}}{\text{Ult Normal-Theory Var (t)}} = \frac{r + R}{1 + rR}$$

i.e., the same expression as that obtained for the ratio of standard deviations of $F = t^2$. However, unlike t^2, the ultimate expected value of t is conveniently equal to the ultimate normal-theory expected value, which is zero. The expected value of ultimate t, i.e., of n'/d', is 0/0 which is indeterminate. However, the expected value of $n' = \bar{P}-\bar{Q}$ is zero at all values of N while the corresponding denominator is a positive nonzero value at all finite values of N and d' becomes zero only when N is infinite. (One may regard d' as a very close approximation to the actual denominator of t at large finite values of N.) Therefore, the indeterminate 0/0 obtained for the ultimate E (t) must actually represent zero. It then follows that

$$\text{Ult } E(t) = \text{Ult Normal-Theory } E(t)$$

so that the ratio of Var (t) to Normal-Theory Var (t), found above, is all that is needed for a serviceable index of the ultimate robustness of t. The index, however, is rendered somewhat more concretely interpretable by taking the square root of the ratio of variances, thereby obtaining an index in terms of standard deviations of actual vs. normal-theory t distributions,

$$\frac{\text{Ult Standard Deviation } (t)}{\text{Ult Normal-Theory Standard Deviation } (t)} = \sqrt{\frac{r + R}{rR + 1}}$$

This ratio is plotted as ordinate against $\sqrt{R}$, i.e., the ratio between the standard deviations of the two populations, as abscissa in Figures 61 - 63 for various constant values of r. It is clear both from the curves and from the formula upon which they are based that there is no unqualified upper bound for the ratio of ultimate standard deviations of t. At $R = 0$ the ratio becomes $\sqrt{r}$ which has no finite upper bound, and at $R = \infty$ the ratio becomes $\sqrt{1/r}$ which also has no finite upper bound (given that r can be less than 1 as well as greater). In both cases the lower bound is zero which, in the present context, is just as drastic as the upper bound of infinity. (Likewise at $r = 0$ the ratio becomes $\sqrt{R}$, at $r = \infty$ it becomes $\sqrt{1/R}$, and these values also have upper and lower bounds of infinity and zero.)

Since it is entirely arbitrary which population is labeled population P, $\sqrt{\frac{r + R}{rR + 1}}$ yields the same value when $R = \sigma_Q^2/\sigma_P^2$ and $r = N_Q/N_P$ as when $R = \sigma_P^2/\sigma_Q^2$ and $r = N_P/N_Q$, i.e., yields the same value when their reciprocals are substituted for R and r. Thus the curves for $R = \Delta$, $r = \theta$ should have essentially the same appearance, e.g., should be mirror images of those for $R = 1/\Delta$, $r = 1/\theta$. This requirement was met, as in the case of the F curves, by using mutually reciprocal abscissa scales on opposite sides of $R = 1$, one of which was always a reciprocal scale and the other, a linear, i.e., reciprocal-of-reciprocal, scale. However, because the labeling of populations is entirely arbitrary, the population yielding the larger number of observations may be labeled population Q, i.e., the population whose variance is $R\ \sigma^2$ and which yields r N observations, thus forcing r to be ≥ 1. And there is, therefore, actually no need to present any curves for $r < 1$, so none is given. Finally if the reciprocal of either R or r (but not both) is substituted in $\sqrt{\frac{r + R}{rR + 1}}$, this ultimate ratio of t standard deviations becomes the reciprocal of the value it had before substitution. As a consequence of this fact and of the scales used, the fourth quadrant of the graphs in Figures 61 and 62 is simply a "180° rotation" of the second quadrant, and in terms of "information" yielded a single quadrant is all that is actually required (see figure 63).

Figures 61-63 show the ultimate robustness of t in terms of the ratio between ultimate standard deviations of actual and normal-theory t. The figures are therefore comparable to those already presented for the ultimate robustness of F. In the case of the F statistic, the ratio between standard deviations was about as direct an index of robustness as we could easily obtain because, under heterogeneity, ultimate F is distributed (in the two population case) as the weighted sum of three chi-square variates. Ultimate t, on the other hand, is normally distributed (since ultimate t is the ratio of a normally distributed variate, $\bar{P}$-$\bar{Q}$, to a constant denominator) irrespective of the presence or absence of heterogeneity, and we can therefore use that additional information to obtain the actual ratio between ρ and α.

Under normal theory, ultimate t is normally distributed with zero mean and unit variance, i.e., it is a standardized normal deviate. Therefore, letting $N(\theta, \lambda)$ stand for a normally distributed variate with mean θ and variance λ (more precisely letting it stand for the product of the normal variate's density function and derivative) and letting Z_α stand for that value of the standardized normal deviate corresponding to a right-tailed cumulative probability, i.e., significance level, of α, we have in the normal-theory case where all assumptions are met

$$\rho = \alpha = \int_{Z_\alpha}^{\infty} N(0, 1) .$$

SENSITIVITY OF t TO HETEROGENEITY OF VARIANCE WHEN SAMPLE SIZE IS INFINITE

RATIO OF ULTIMATE (N=∞) STANDARD DEVIATION OF t DISTRIBUTION TO ULTIMATE NORMAL-THEORY (R=1) STANDARD DEVIATION OF t, WHEN t STATISTIC IS BASED UPON SAMPLES OF SIZE N AND rN, RESPECTIVELY, FROM NORMAL POPULATIONS WITH EQUAL MEANS BUT VARIANCES σ^2 AND $R\sigma^2$

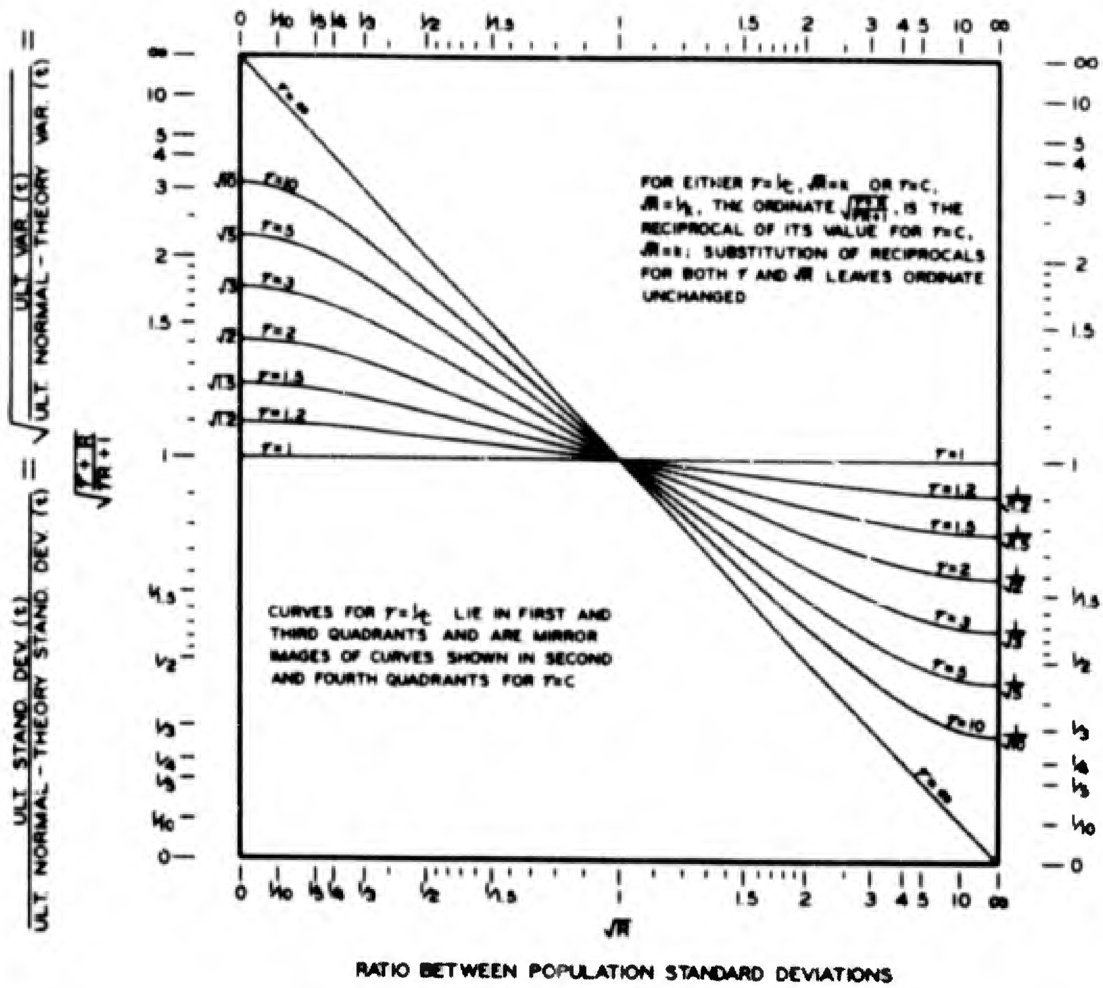

Figure 61. Ultimate Robustness of t Statistic

SENSITIVITY OF t TO HETEROGENEITY OF VARIANCE WHEN SAMPLE SIZE IS INFINITE

RATIO OF ULTIMATE ($N=\infty$) STANDARD DEVIATION OF t DISTRIBUTION TO ULTIMATE NORMAL-THEORY ($R=1$) STANDARD DEVIATION OF t, WHEN t STATISTIC IS BASED UPON SAMPLES OF SIZE N AND rN, RESPECTIVELY, FROM NORMAL POPULATIONS WITH EQUAL MEANS BUT VARIANCES σ^2 AND $R\sigma^2$

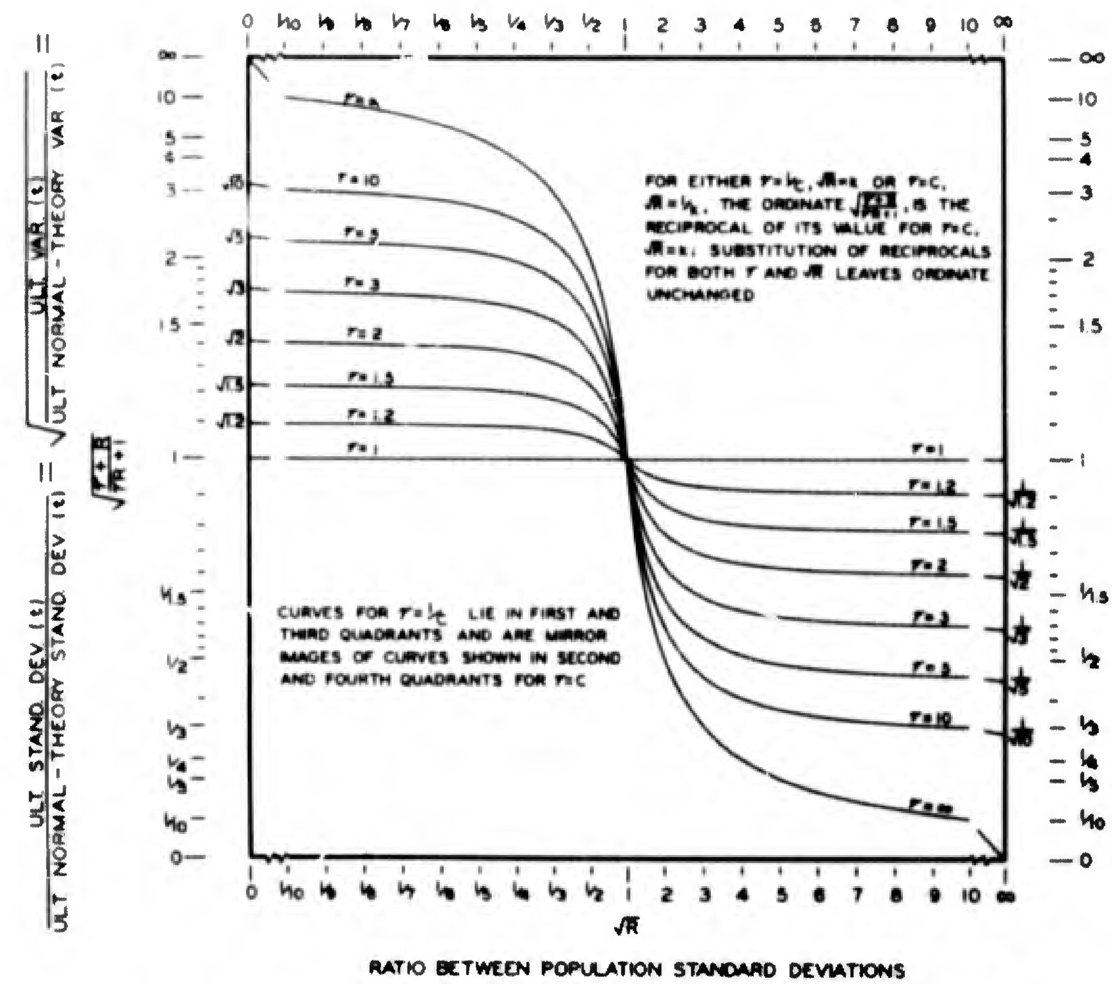

Figure 62. Same Data as Figure 61, Plotted to Different Abscissa Scale

SENSITIVITY OF t TO HETEROGENEITY OF VARIANCE WHEN SAMPLE SIZE IS INFINITE

RATIO OF ULTIMATE ($N=\infty$) STANDARD DEVIATION OF t DISTRIBUTION TO ULTIMATE NORMAL-THEORY ($R=1$) STANDARD DEVIATION OF t, WHEN t STATISTIC IS BASED UPON SAMPLES OF SIZE N AND τN, RESPECTIVELY, FROM NORMAL POPULATIONS WITH EQUAL MEANS BUT VARIANCES σ^2 AND $R\sigma^2$

(NOTE: ORDINATE VALUES FOR $\tau = C$ ARE RECIPROCALS OF THOSE FOR $\tau = 1/C$; FOR A GIVEN τ ORDINATE VALUES FOR $\sqrt{R} = 1/k$ ARE RECIPROCALS OF THOSE FOR $\sqrt{R} = k$)

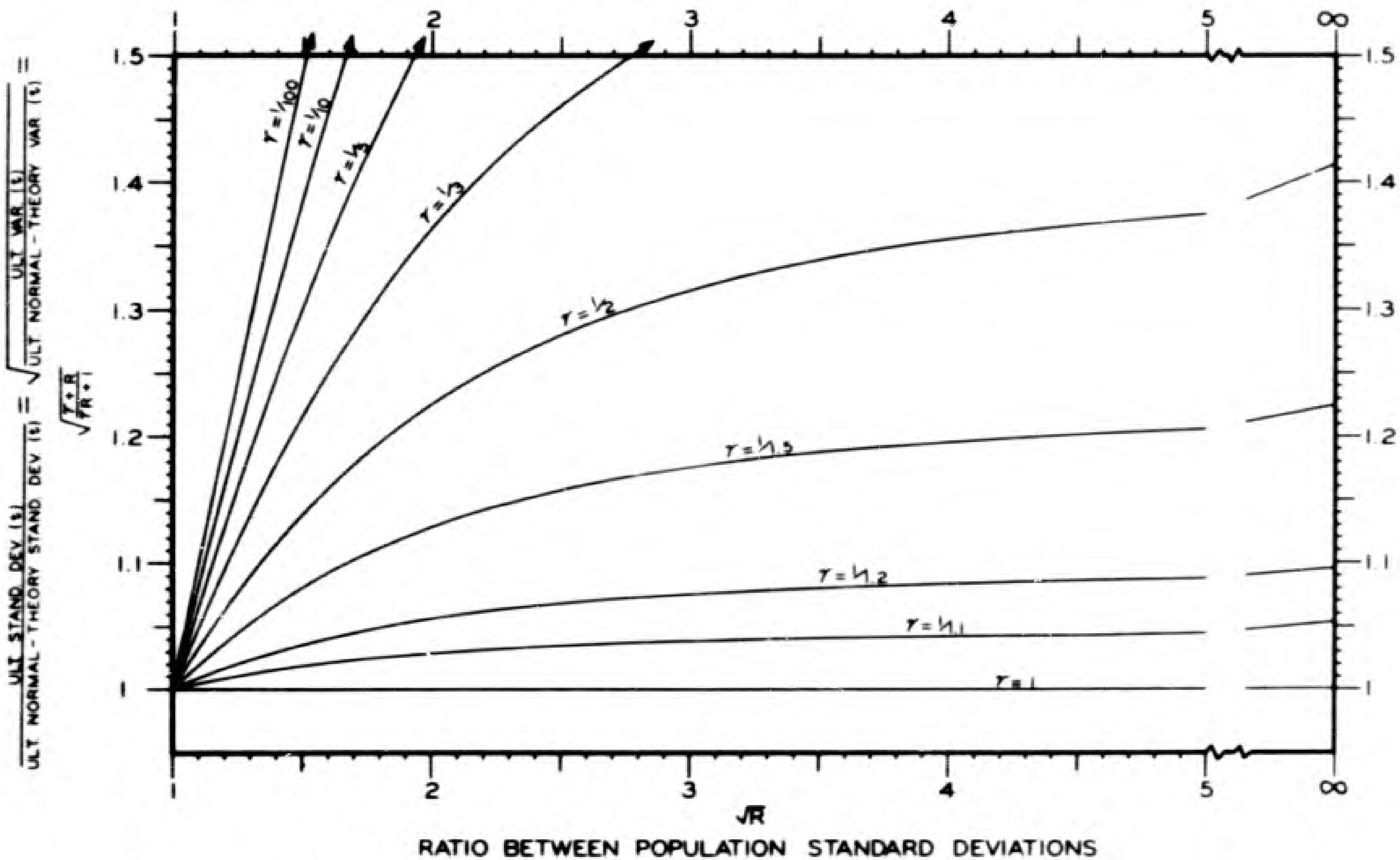

Figure 63. Critical Section of (unplotted first quadrant of) Figure 62, Plotted to Linear Ordinate Scale

Under heterogeneity (with or without nonnormality) however, ultimate t is also a normally distributed variate, and it also has zero mean, but has variance of (r + R)/(rR + 1). So,

$$\rho = \int_{Z_\alpha}^{\infty} N\left(0, \frac{r + R}{rR + 1}\right),$$

or, we can standardize the normal distribution indicated above (for ultimate t under heterogeneity) by dividing all values in it by its standard deviation, $\sqrt{(r + R)/(rR + 1)}$, in which case

$$\rho = \int_{\frac{Z_\alpha}{\sqrt{\frac{r + R}{rR + 1}}}}^{\infty} N(0, 1)$$

Thus ρ is simply the right-tailed cumulative probability of $Z_\alpha/\sqrt{(r + R)/(rR + 1)}$ in the standard normal distribution. It can therefore be obtained directly from tables of the cumulative normal distribution or, more accurately, by integration using an electronic computer. The former method was used to obtain the data for Figure 64, which gives the ratio between ρ and α as a function of $\sqrt{(r + R)/(rR + 1)}$. Since the latter is what is plotted as ordinate in Figures 61-63, Figure 64 provides the means of "translating" those ordinate values (i.e., ratios between standard deviations of nonnormal-theory and normal-theory t) into ratios between true and normal-theory significance levels. The curves are steep, indicating that heterogeneity has an even more drastic effect upon standard significance levels (i.e., upon the probability of a Type I Error) than upon the standard deviation of the distribution of the t statistic. Furthermore, the situation worsens rapidly as α diminishes. Computer integration was used to obtain the data for Figures 65-70 which obviate the "translation" by plotting the ratio between ρ and α as ordinate (instead of the less directly meaningful ratio between standard deviations of t) against $\sqrt{R}$ as abscissa. Because of the symmetry of the normal distribution, the ratio between ρ and α is the same for a left-tailed as for the corresponding right-tailed α and the ratio for a two-tailed α of .05, .01, or .001 is the same as

RATIO BETWEEN TRUE, ρ, AND NORMAL-THEORY, α, ONE-TAILED SIGNIFICANCE LEVELS AS A FUNCTION OF THE RATIO BETWEEN TRUE AND NORMAL-THEORY STANDARD DEVIATIONS OF THE t DISTRIBUTION (WHEN t STATISTIC IS BASED UPON INFINITE-SIZED SAMPLES CONSISTING OF N AND γN OBSERVATIONS, RESPECTIVELY, FROM NORMAL POPULATIONS WITH EQUAL MEANS BUT VARIANCES σ^2 AND $R\sigma^2$)

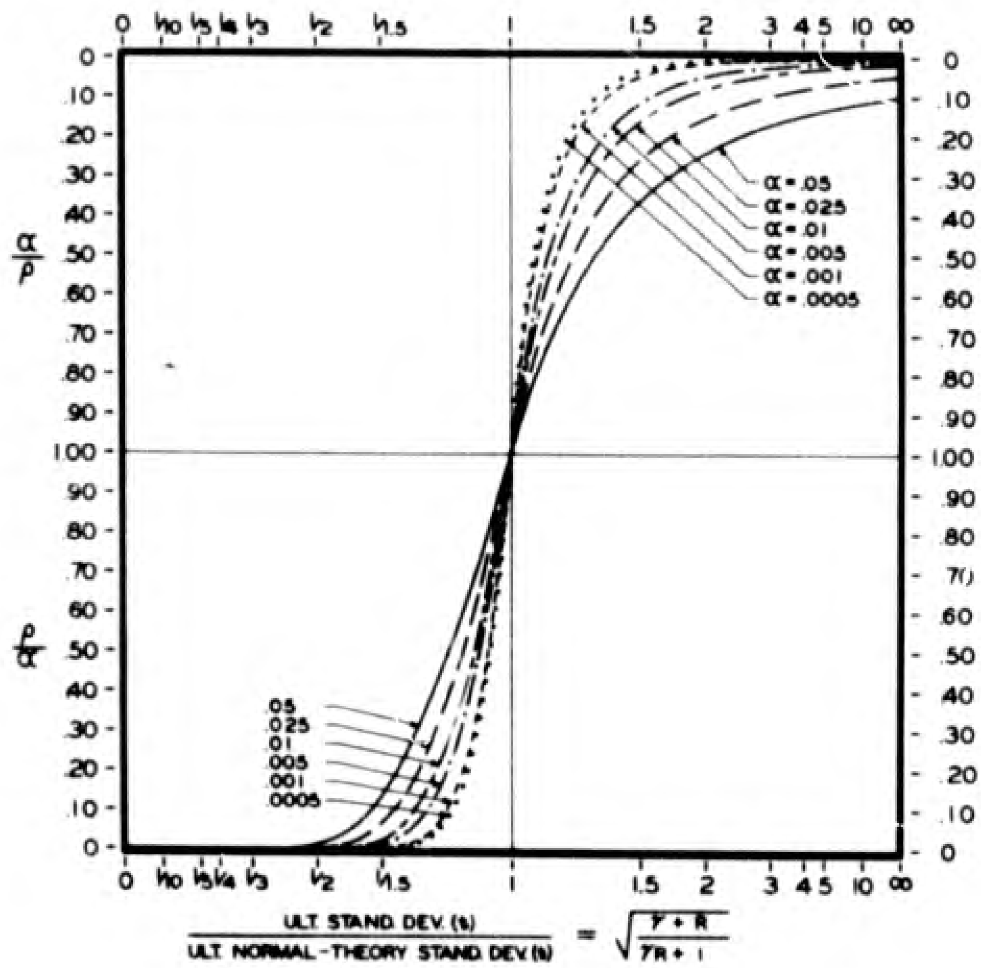

Figure 64. Relationship between $\frac{\text{Ult Stand. Dev. (t)}}{\text{Ult Normal-Theory Stand. Dev. (t)}}$ (i.e., ordinates of figures 61-63) and ρ/α, (i.e., ratio between true and normal-theory significance levels)

that for a one-tailed α of .025, .005, or .0005 respectively. Therefore, the three standard significance levels of .05, .01, and .001 are covered for both one-tailed and two-tailed tests by investigating the six one-tailed α's of .05, .025, .01, .005, .001, and .0005. Each of the figures numbered 65 to 70 give the relationship between ρ/α or α/ρ as ordinate and $\sqrt{R}$, as abscissa for one of these six conditions.

The points whose abscissae are $\sqrt{R} = .5$ and $\sqrt{R} = 2$ on the curve labeled r = 2 and on the curve labeled r = 3, respectively, correspond to the empirically investigated statistics TYB N2N, TYB 2NN, TYB N3N, and TYB 3NN for the limiting case where $N = \infty$. (And, of course, TYB 2N2N at $N = \infty$ corresponds to the points having an abscissa of either $\sqrt{R} = .5$ or $\sqrt{R} = 2$ on the curve labeled r = 1, a curve which is a straight horizontal line representing a condition of complete ultimate robustness.) In fact the above statements also hold if for YB we substitute XA, XB, or YA, i.e., the statement holds for those statistics violating the normality assumption in addition to that of heterogeneity. However, while in the YB cases the terminal values of the empirical ρ/α or α/ρ ratios at N = 1024 correspond closely with the theoretical values for $N = \infty$, this is not always true for the statistics also violating the normality assumption. As a matter of empirical fact the effects of heterogeneity alone, which persist all the way to $N = \infty$, appear to have become stabilized at very close to their limiting or asymptotic values when sample size is quite small, e.g., N = 16, while contrastingly the combined effects of nonnormality and the nonnormality-heterogeneity interaction, which become nil at $N = \infty$, often have not yet washed out at N = 1024.

Mathematically derived results for the F statistic in general are also supported by the empirical results of the sampling study. When only the assumption of normality was violated the empirical curves (showing the ratio between ρ and α for FXX and FXY) appeared, above N = 2, to approach a value of 1.00 as an asymptote (except for fluctuations ascribable to chance). (At N = 2 discreteness effects are large and presumably account for anomolous behavior of the FXX curves in that region.) And at some point prior to N = 1024 the curves had always begun to fluctuate about the horizontal line through 1.00 in an essentially random fashion (i.e., the contribution of assumption-violation to deviations of the curves from 1.00 appeared to have become negligible relative

SENSITIVITY OF t TEST TO HETEROGENEITY OF VARIANCE WHEN BOTH SAMPLES ARE OF INFINITE SIZE

RATIO BETWEEN TRUE, ρ AND NORMAL-THEORY, α, PROBABILITIES OF REJECTION WHEN TEST USES ONE-TAILED REJECTION REGION CORRESPONDING TO α = .05 AND t STATISTIC IS BASED UPON SAMPLES OF SIZE N AND γN, RESPECTIVELY, FROM NORMAL POPULATIONS WITH EQUAL MEANS BUT VARIANCES σ^2 AND $R\sigma^2$

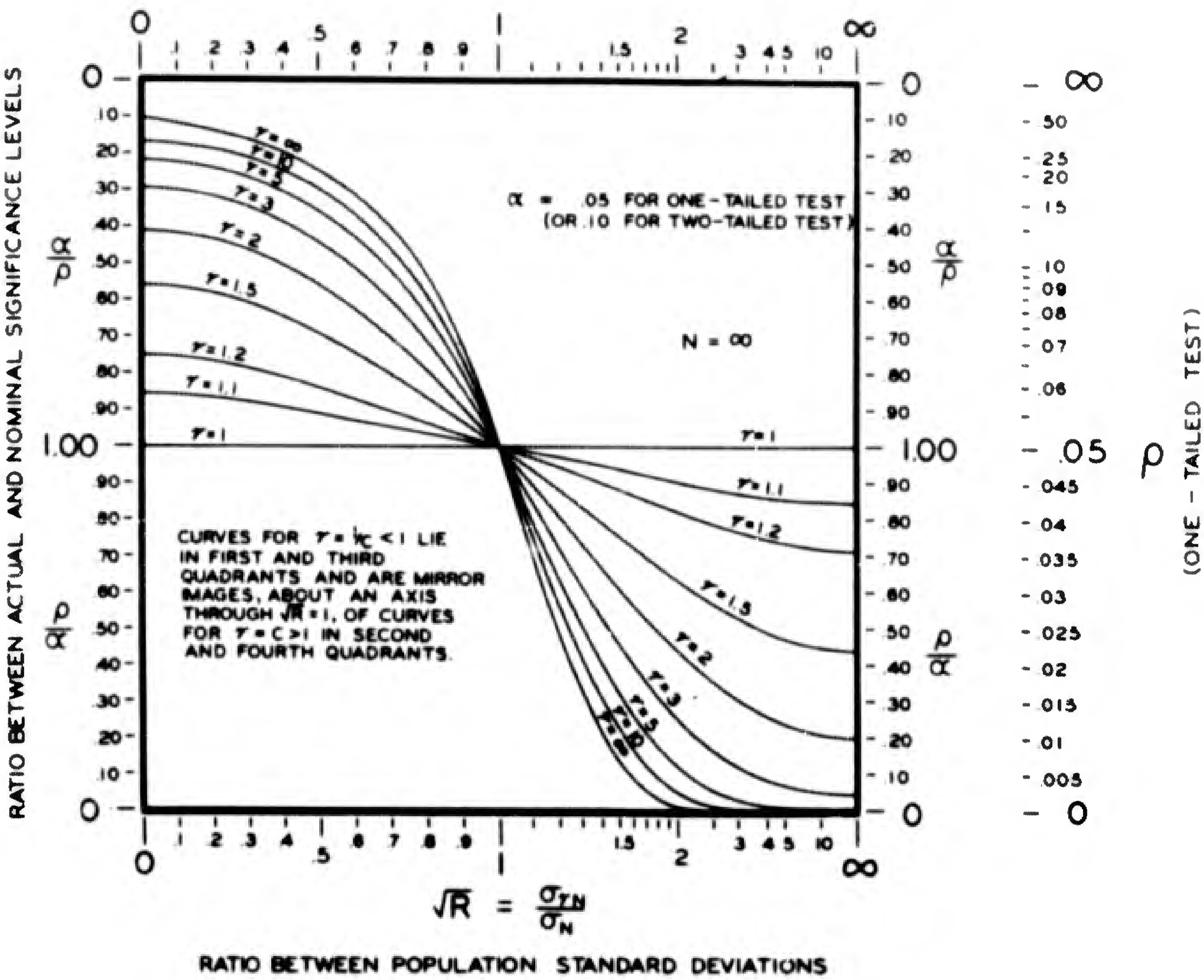

Figure 65. Ultimate Robustness of One-Tailed t Test Conducted at .05 Level of Significance

SENSITIVITY OF t TEST TO HETEROGENEITY OF VARIANCE WHEN BOTH SAMPLES ARE OF INFINITE SIZE

RATIO BETWEEN TRUE, ρ AND NORMAL-THEORY, α, PROBABILITIES OF REJECTION WHEN TEST USES TWO-TAILED REJECTION REGION CORRESPONDING TO $\alpha = .05$ AND t STATISTIC IS BASED UPON SAMPLES OF SIZE N AND γ N, RESPECTIVELY, FROM NORMAL POPULATIONS WITH EQUAL MEANS BUT VARIANCES σ^2 AND $R\sigma^2$

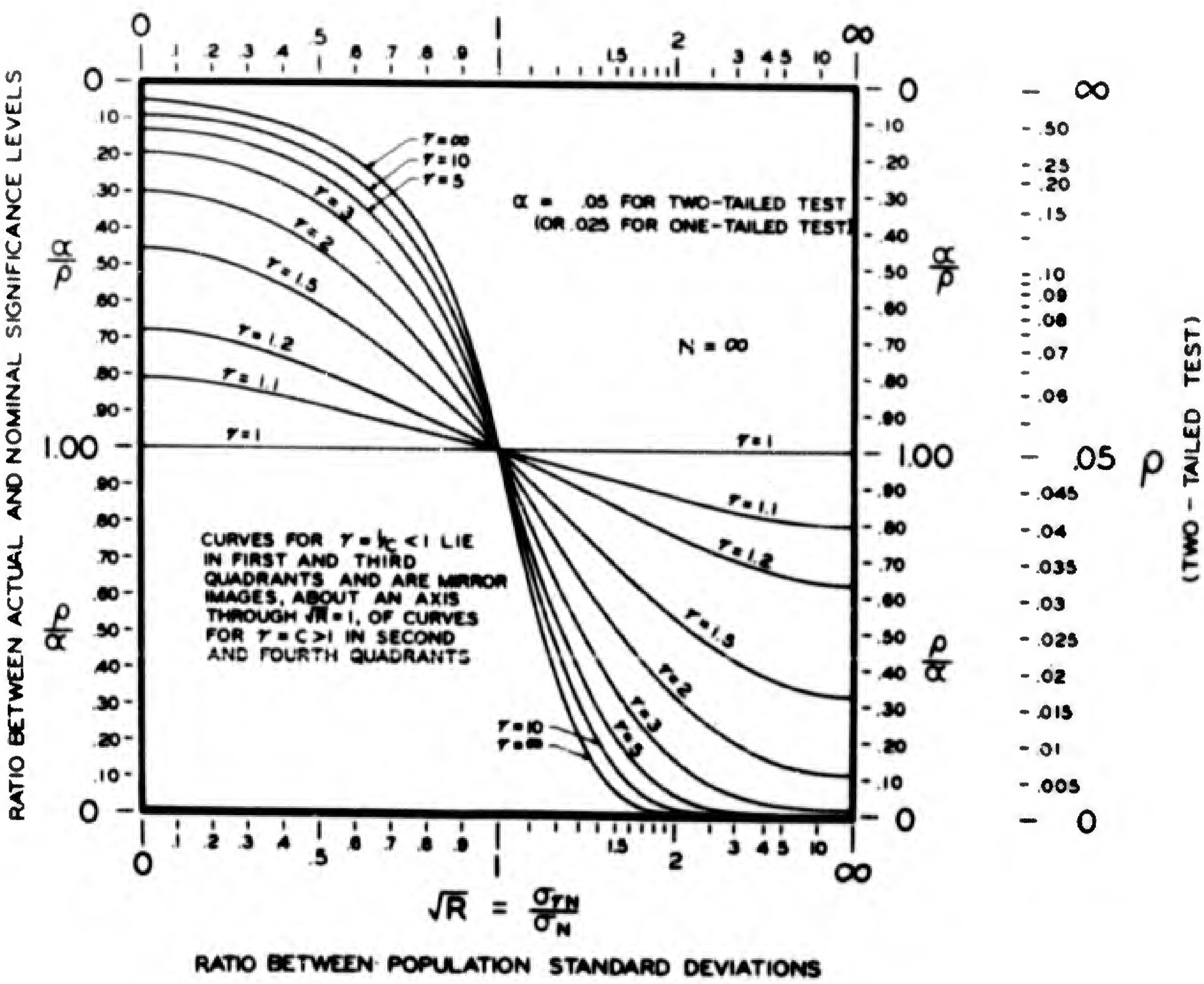

Figure 66. Ultimate Robustness of Two-Tailed t Test Conducted at .05 Level of Significance

SENSITIVITY OF t TEST TO HETEROGENEITY OF VARIANCE WHEN BOTH SAMPLES ARE OF INFINITE SIZE

RATIO BETWEEN TRUE, ρ, AND NORMAL-THEORY, α, PROBABILITIES OF REJECTION WHEN TEST USES ONE-TAILED REJECTION REGION CORRESPONDING TO $\alpha = .01$ AND t STATISTIC IS BASED UPON SAMPLES OF SIZE N AND γN, RESPECTIVELY, FROM NORMAL POPULATIONS WITH EQUAL MEANS BUT VARIANCES σ^2 AND $R\sigma^2$

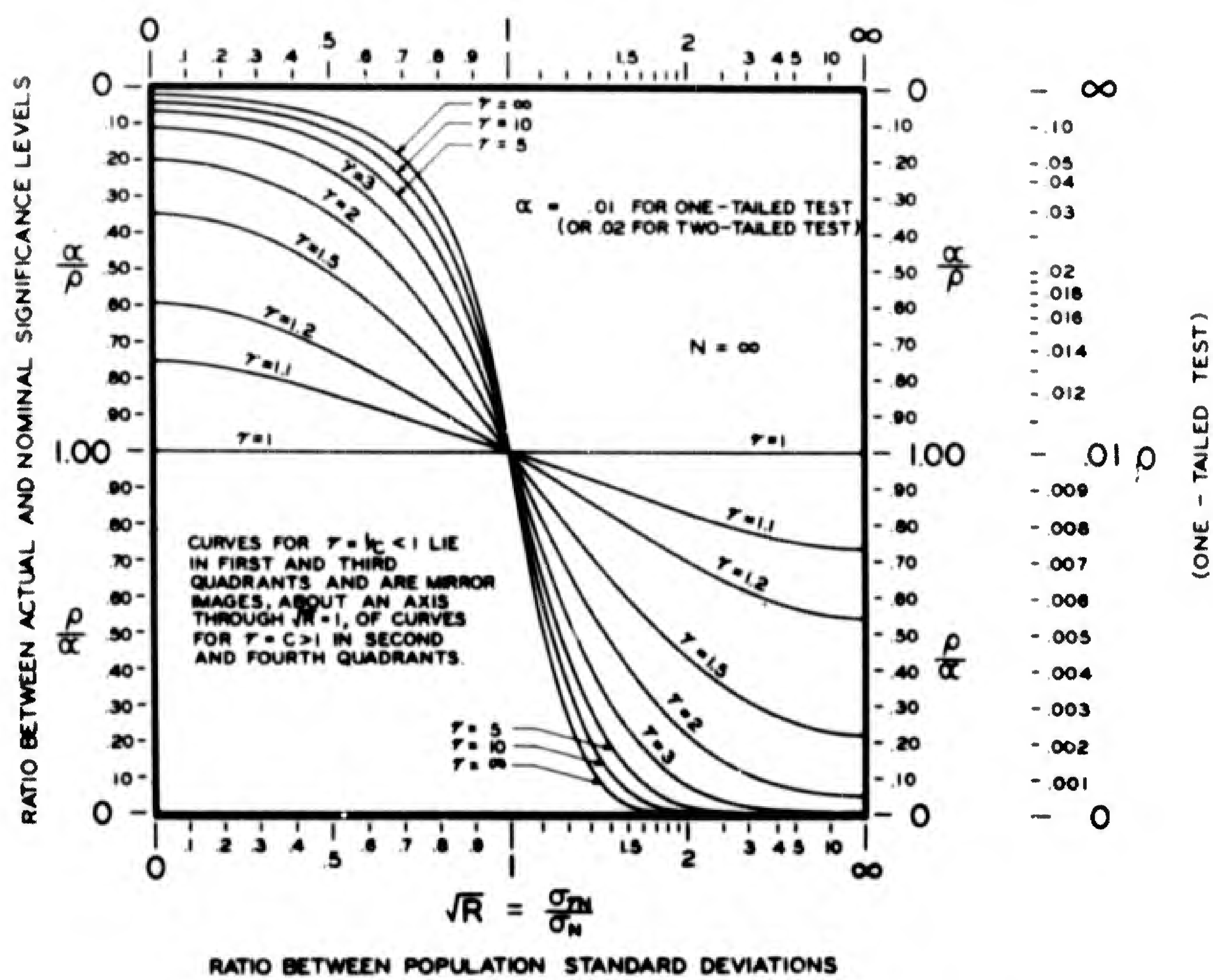

Figure 67. Ultimate Robustness of One-Tailed t Test Conducted at .01 Level of Significance

SENSITIVITY OF t TEST TO HETEROGENEITY OF VARIANCE WHEN BOTH SAMPLES ARE OF INFINITE SIZE

RATIO BETWEEN TRUE, ρ AND NORMAL-THEORY, α, PROBABILITIES OF REJECTION WHEN TEST USES TWO-TAILED REJECTION REGION CORRESPONDING TO $\alpha = .01$ AND t STATISTIC IS BASED UPON SAMPLES OF SIZE N AND γ N, RESPECTIVELY, FROM NORMAL POPULATIONS WITH EQUAL MEANS BUT VARIANCES σ^2 AND $R\sigma^2$

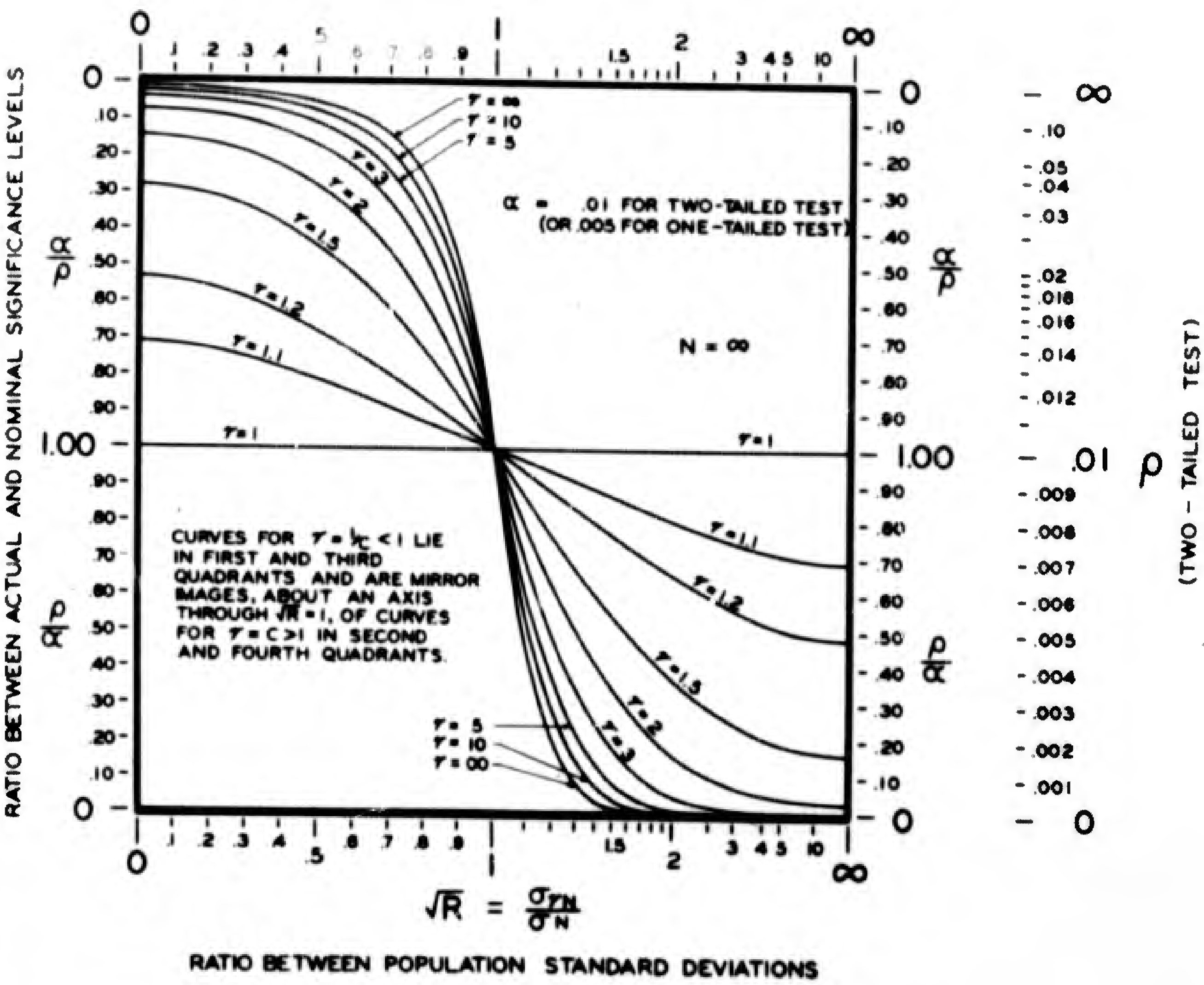

Figure 68. Ultimate Robustness of Two-Tailed t Test Conducted at .01 Level of Significance

SENSITIVITY OF t TEST TO HETEROGENEITY OF VARIANCE WHEN BOTH SAMPLES ARE OF INFINITE SIZE

RATIO BETWEEN TRUE, ρ, AND NORMAL-THEORY, α, PROBABILITIES OF REJECTION WHEN TEST USES ONE-TAILED REJECTION REGION CORRESPONDING TO $\alpha = .001$ AND t STATISTIC IS BASED UPON SAMPLES OF SIZE N AND γN, RESPECTIVELY, FROM NORMAL POPULATIONS WITH EQUAL MEANS BUT VARIANCES σ^2 AND $R\sigma^2$

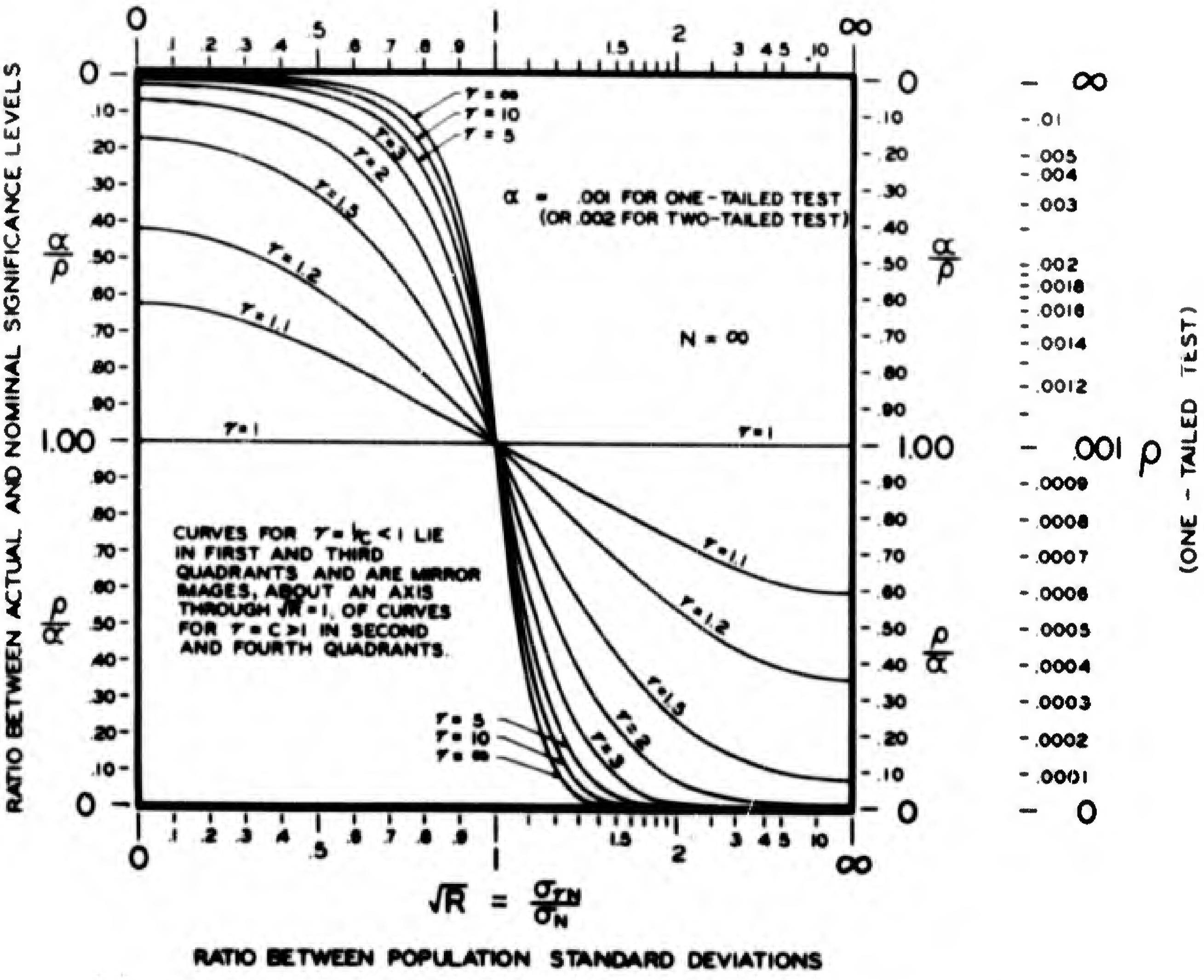

Figure 69. Ultimate Robustness of One-Tailed t Test Conducted at .001 Level of Significance

SENSITIVITY OF t TEST TO HETEROGENEITY OF VARIANCE WHEN BOTH SAMPLES ARE OF INFINITE SIZE

RATIO BETWEEN TRUE, ρ, AND NORMAL-THEORY, α, PROBABILITIES OF REJECTION WHEN TEST USES TWO-TAILED REJECTION REGION CORRESPONDING TO $\alpha = .001$ AND t STATISTIC IS BASED UPON SAMPLES OF SIZE N AND τN, RESPECTIVELY, FROM NORMAL POPULATIONS WITH EQUAL MEANS BUT VARIANCES σ^2 AND $R\sigma^2$

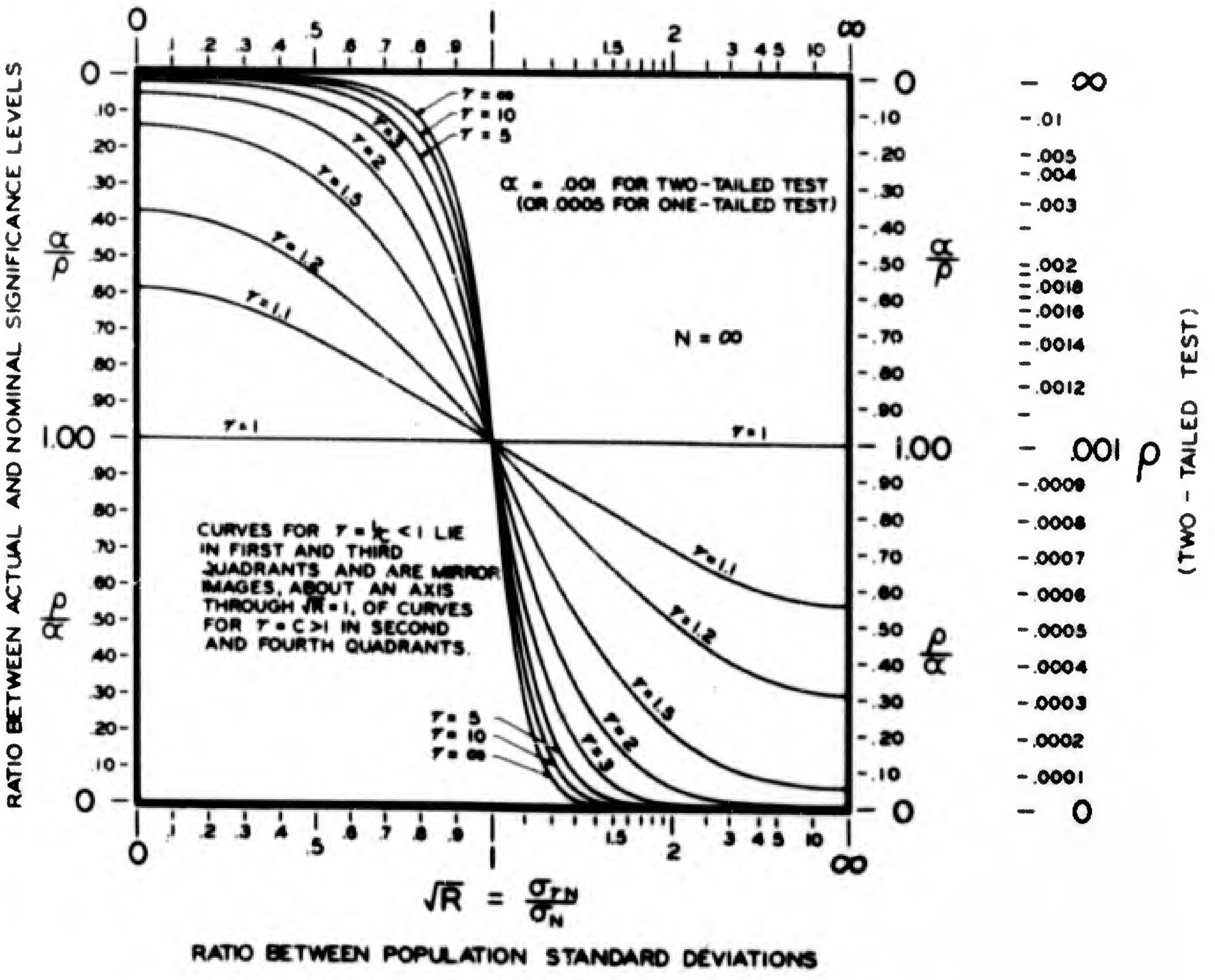

Figure 70. Ultimate Robustness of Two-Tailed t Test Conducted at .001 Level of Significance

to the contribution of "chance" to them). A similar statement holds for curves for the two-tailed TXX and TXY tests, i.e., for two-sample F tests under nonnormality alone. These empirical results support the mathematically based conclusion that F is ultimately completely robust against violation of the normality assumption alone. In every case of heterogeneity of variance, whether accompanied by nonnormality or not, all of the empirical, multi-sample F curves lay entirely above the horizontal line through 1.00, i.e., lay in the region where $\alpha/\rho < 1$ and therefore where ρ exceeds α, and appeared to be approaching limiting values different from 1.00. These facts are in harmony with the mathematical findings that, when based upon unequal numbers of equal-sized samples drawn from two populations with the same mean but unequal variances, (a) the F test is not ultimately completely robust and (b) the standard deviation of the F distribution can only be larger, never smaller, than the normal-theory standard deviation. In cases of heterogeneity of variance and unequal sample sizes, the empirical curves for two-tailed T's, at N = 1024, indicate that ρ exceeds, or is exceeded by, α according to whether the ultimate standard deviation of t exceeds, or is exceeded by, the ultimate normal-theory standard deviation of t, respectively. Finally, there is a perfect rank order correlation between the ratio of ρ to α at N = 1024 and the ratio of ultimate standard deviation to ultimate normal-theory standard deviation for all of the FYB and two-tailed TYB statistics, at any of the standard values of α, i.e., at any value of α appearing in the empirical graphs.

Having explored the ultimate robustness of F, we now consider its behavior when N is finite and particularly when it is small enough for the shape and variance of the denominator's distribution to exert an appreciable influence upon the robustness of F. Consider first the case of violation of the normality assumption alone. When N is very small, there is a substantial probability that all N observations in a sample will be drawn from the tightly concentrated hump surrounding the primary mode of the X population. When this happens, $\sum_1^N (X-\bar{X})^2$ is very small, relative to normal-theory expectation, and, because the primary mode lies just below u, $\bar{X}$ is slightly smaller than u. If it happens for all k samples, $\sum_1^k \sum_1^N (X-\bar{X})^2$ will be very small relative to normal-theory expectation and all the $\bar{X}$'s will lie within a very narrow range of values just below μ, so that $\sum_1^k (\bar{X}-\bar{\bar{X}})^2$ will also be very small. The result is that, when the X population is the only population sampled, very small sums of squares in the denominator of F are associated with very small sums of squares in the numerator. Observations which are not drawn from the hump of distribution surrounding the primary mode must come from the long, thin, positive skew which extends for 10 σ above μ. If a small to moderate-sized sample contains even a very small number of skew observations, its sum of squares tends to be excessively large relative to normal-theory expectation and its mean tends to lie above μ. If each of k samples, or a fairly large proportion of them, contain some skew observations, $\sum_1^k \sum_1^N (X-\bar{X})^2$ will tend to be excessively large relative to normal theory and the $\bar{X}$'s will tend to be spread out over a wide range of values above μ, so that $\sum_1^k (\bar{X}-\bar{\bar{X}})^2$ will also be excessively large. Thus large sums of squares in the denominator are associated with large sums of squares in the numerator.

Sums of squares are notoriously sensitive to nonnormality. It has already been implied above that $\sum_1^N (X-\bar{X})^2$ assumes, at far greater frequency than would be expected under normal theory, values which are both extremely large and extremely small relative to normal-theory expectation. It is shown in figure 53 that this is indeed the case and, furthermore, that the extreme nonrobustness of $\sum_1^N (X-\bar{X})^2$ does not improve appreciably with increasing N. From the considerations of the preceding paragraph (and from the fact that the distribution of $\bar{X}$ progresses from identity with the X distribution at $N = 1$ to an eventual normal distribution at $N = \infty$) it is reasonable to suppose that the sum of squares in the numerator, $\sum_1^k (\bar{X}-\bar{\bar{X}})^2$, also exceeds normal-theory frequency in assuming normal-theory "tail values" at both tails and that this tendency is undiminished by increasing k, but is greatly reduced with increasing N, due to the reduction of $\bar{X}$'s nonnormality by the Central Limit effect.

Because of the positive correlation between numerator and denominator, the effect upon F of the tendency for the denominator to assume extremely large (or extremely small) values is offset and counteracted by a similar tendency in the numerator so that the F ratio tends to remain moderate. In fact the correlation is so strong that there is actually a deficit of upper tail F values relative to normal-theory expectation (i.e., $\rho < \alpha$ at nonrobust N's for FXX statistics, at least above $N = 2$).

However, if even a single sample is drawn from the Y population, the effect is that of introducing some random "noise" which weakens the correlation between total numerator and total denominator. (Means and variances of samples drawn from normal populations are uncorrelated.) With the breaking of their fetters, large numerators can occur together with moderate or small denominators, and small denominators can occur together with moderate or large numerators, and both concurrences produce large F ratios. The result is that ρ exceeds α for the FXY statistics at nonrobust N values. Since the denominator is a sort of weighted average of sums of squares, the weight given the X observations is reduced in accordance with the proportion of observations which are Y observations, the effect of which is to tend to increase robustness. However, the greater the proportion of Y observations, the greater the attenuation of the correlation between numerator and denominator of F, and the effect of this attenuation is to tend to increase nonrobustness. Thus, one effect tends to offset the other: FXY N,NNN is only very slightly more robust than FXY N,NN, and FXY NNN,N, rather than being less robust than FXY NN,N, appears to be somewhat more so. The latter effect and the fact that FXX NNNN is somewhat more robust than FXX NNN are presumably due to the same cause: The larger the number of equal-sized samples from the same nonnormal population, the greater the opportunity for extremely large and extremely small sample variances to cancel each other out or be otherwise attenuated in the averaging process which takes place in the denominator of F.

As N increases, the means in the numerator of F violate the normality assumption less and less seriously, due to the Central Limit effect, and the numerator becomes increasingly robust. The denominator remains drastically nonrobust (at least in regard to the shape and variance of its distribution), showing no appreciable improvement with increasing N, but that nonrobustness becomes increasingly irrelevant due to the shrinking variance of the denominator about the parameter it estimates. The ultimate result, as already indicated, is that as N approaches infinity, F approaches complete robustness and $\bar{\rho}$ approaches a limiting value of α.

Now consider the case of violation of the assumption of homogeneity of variance alone. Other things being equal, when the population with the smaller variance contributes a single sample, the variance of that sample's observations should tend to fall well within the "expected" range of variances for the other samples, but when the population with the larger variance contributes a single sample, that sample's variance should frequently fall outside of say a 95% tolerance region for variances for the other samples. In the former case the "exceptional" population produces an effect very similar to what would have been expected had an additional sample been drawn, instead, from the nonexceptional population. In the latter case, however, the effect is quite dissimilar. Consequently one would expect the robustness of the denominator to suffer more in the latter case than in the former. The situation in the numerator is analogous. Other things being equal (such as skewedness), the larger the relative variance of the population contributing the single sample, the greater the likelihood that that sample mean will fall outside the range of values "expected" for the sample means from the other population, i.e., outside of a reasonable tolerance region for means of samples from the other population. Hence the greater the detriment to the robustness of the variance estimate in the numerator. Since both numerator and denominator should be more robust when the exceptional population has the smaller variance, one would expect F to be more robust also, and the empirical data confirm that it is (with rare exceptions; see FXB at small N's).

Naively, one might suppose that the larger the number of samples drawn from the nonexceptional population the smaller would be the effect of the single sample from the exceptional population upon the robustness of F. However, Figures 58-60, and the derivations upon which they are based, suggest that this is the case only after the ratio between the numbers of samples drawn from the two populations has passed a certain critical point (which depends upon both the relative variance of the exceptional population and the absolute number of samples drawn from it) and that the opposite is the case before this critical point is reached. When the nonexceptional population has the larger variance, the critical point is low enough to give considerable support to the naive supposition but this is no longer the case when the nonexceptional population has the smaller variance. In any event, Figures 58-60 suggest (on the basis of results at infinite N) that FYB NNN,N should be very slightly more robust than FYB NN,N and that FYB N,NNN should be considerably less robust than FYB N,NN. The latter was in fact the case; the former was the case at N = 1024, but not in general, a slight difference in the opposite direction being perhaps ascribable to "chance".

No attempt will be made to explain the combined effects of two violations of assumptions. However, one paradoxical interaction effect is worthy of note. The effect of nonnormality alone in all sampled populations was to render ρ less than α (at moderate N's above N = 2. See FXX graphs). The effect of heterogeneity of variance alone was to render ρ greater than α (see FYB graphs). However when both nonnormality in all sampled populations and heterogeneity of variance were present, their combined effect, rather than being a sort of average in which $\rho < \alpha$ and $\rho > \alpha$ tend to offset one another, was to render $\rho >> \alpha$, i.e., to make ρ exceed α still more than was the case under heterogeneity alone (see FXA graphs).

It has already been pointed out that the two-tailed, two-sample t test is equivalent to a two-sample F test. Therefore when used as a two-tailed test, the statistic TPQ C_1NC_2N might just as validly be designated FPQ C_1N,C_2N. As such it is especially comparable with the FPQ statistic based upon the same total number of observations drawn from each of the same two populations but combining the larger number of observations into C samples each of size N rather than into a single sample of size CN. Thus when used as a two-tailed test, each of the following T statistics differs from the F statistic with which it is paired only in the number of samples into which the observations have been combined: TPQ N3N and FPQ N,NNN; TPQ N2N and FPQ N,NN; TPQ 2NN and FPQ NN,N; TPQ 3NN and FPQ NNN,N. So by comparing the graph for the two-tailed case of TPQ 2NN with the graph for FPQ NN,N we are comparing the robustness of FPQ when the 2N observations from the P population are treated as a single sample with the robustness of FPQ when the 2N observations from the P population are treated as two equal-sized but separate samples. If we make all of the appropriate comparisons, we find that for all cases involving heterogeneity of variance the F statistic is more robust (with highly localized exceptions) than the corresponding two-tailed T statistic, but when variances are homogeneous either statistic may be the more robust. (This finding is analogous to the discovery that the Z test was always more robust than the T test when variances were heterogeneous and sometimes, but not always, so when variances were homogeneous.) TXX 2N,N, TXX 3NN, and TXY N2N were more robust than FXX NNN, FXX NNNN and FXY N,NN, respectively. In other cases characterized by homogeneity of variance, the F test was the more robust. Another comparison justified by the above reasoning is that of two-tailed TXX 3NN with FXX 2NNN and FXX 2NNN with FXX NNNN. From most robust to least robust the statistics are TXX 3NN, FXX 2NNN, and FXX NNNN.

The effects mentioned in the preceding paragraph have important practical implications for devotees of the Robustness school. If variances are known or suspected to be heterogeneous and one sample size is an integral multiple, c, of the other (or nearly so), and a two-tailed test is contemplated, robustness may be markedly increased by breaking the larger sample into c samples of equal size and performing an F test rather than a t test. Clearly this procedure discards "information", and under normal-theory conditions this would tend to incur a loss of power. Whether or not the gain in robustness would compensate for the loss of "normal-theory" power, the writer will leave to disciples of the Robustness school to ponder.

To summarize: On the basis of the data and reasoning of the present section as well as of earlier sections it can be concluded that the robustness of the analysis of variance F statistic is influenced, frequently in a complex way involving multiple interactions, by at least the following factors: (a) absolute population shape, i.e., the shape of each sampled population's distribution, (b) relative population shape, i.e., the relationship of each population's shape to that of every other population, (c) the relative variance of each sampled population, (d) the absolute number of samples drawn from each different population and (e) the absolute number of observations contained in each such sample, (f) the relative number of samples drawn from each different population and (g) the relative number of observations contained in each sample, (h) the absolute correlation between mean and variance of samples drawn from a given nonnormal population, (i) the relative amount of each such correlation (for those cases in which either sample size or population shape varies), (j) size of nominal significance level, (k) interactions among factors, e.g., which population shape with which variance yielded how many samples of what sizes. The importance of interactions is illustrated by the fact that when a constant number of samples are drawn from one normal population and all other samples are drawn from a second normal population with the same mean but a different variance (all samples being of equal and infinite size) robustness decreases with the number of samples drawn from the second population up to a critical point, which depends upon the constant number of samples from the first population and upon the relative variance of that population, after which point robustness increases with further increases in the number of samples drawn from the second population. It is clear, therefore, that interaction effects are both important and erratic, even when only a single assumption is violated, and that the various factors may operate in different directions and with strengths which vary with still other factors.

Nonnormality changes the shape and variance, but not the mean of the distribution of F's numerator and denominator, relative to their condition under normal theory. At very small sample sizes the shape and variance of the numerator's distribution are appreciably distorted and this tends to contribute appreciably to the nonrobustness of F. However, the numerator contains only sample means, and as N increases these means become more and more nearly normally distributed, due to the Central Limit effect. Since normally distributed means is all that is actually required in the numerator by the "normality assumption", the numerator becomes increasingly robust as sample size increases and perfectly robust at N = infinity. The shape of the denominator's distribution does not improve appreciably with increasing N. However, the variance of the denominator shrinks rapidly, with increasing N, relative to that of the numerator. As the denominator's distribution shrinks toward its mean, which is unaffected by nonnormality, the robustness of F becomes more and more determined by its increasingly robust numerator. Thus as N increases the F test becomes increasingly robust against a given violation of the normality assumption. Finally, when all sample sizes become infinite, the numerator has exactly its normal-theory distribution and the nonrobustness of the shape of the denominator's distribution has ceased to matter since its variance has become zero relative to that of the numerator, i.e., since the denominator has become functionally a constant, the same constant it would have become under normal theory. Thus at infinite sample sizes the F statistic has exactly its normal-theory distribution and the F test is perfectly robust against violation of the normality assumption.

Heterogeneity of variance tends to shift the means of the distributions of both numerator and denominator of F. When samples are of unequal size the two shifts are unequal and this tends to bias the location of the F distribution relative to its normal-theory location. If the F distribution does not even have its normal-theory location perfect robustness is virtually impossible. When all samples are of the same size, the distributions of the numerator and of the denominator have the same mean so the location of the F distribution remains relatively unaffected. However, the variance of the F distribution, even when sample size is infinite, _is_ distorted by heterogeneity in all practical cases except when $F = t^2$. And the robustness of F is therefore adversely affected. When, under these circumstances, the samples are drawn from only two populations, the distortion is always such that the variance of the F distribution exceeds the variance of the normal-theory F distribution, and consequently ρ exceeds α.

When only the normality assumption was violated the approach of the empirical ρ/α or α/ρ curves toward 1.00 was essentially monotonic (above N = 2). When only the homogeneity assumption was violated, the curves showed little or no trend from N = 2 to N = 1024 and were very nearly straight horizontal lines (except for chance fluctuations) sometimes lying far above 1.00 (depending upon α, the number of samples involved, and the relative variance of the population from which the greater number of samples were drawn). However, when both the normality and homogeneity assumptions were violated, (a) monotonicity was no longer the rule (although nonmonotonicity, when it occurred, was not very drastic), (b) although the two violations individually produced opposite effects upon the size of ρ relative to α, these opposite effects generally did not tend to offset one another or cancel out in interaction; instead, the interaction tended to incorporate the larger of the two distortions (although reversing the direction of the effects of nonnormality!) or even to incorporate both, becoming something analogous to the sum of the absolute values of the separate distortions.

Results of the present study suggest that the F test is less spectacularly, less erratically, and less unpredictably nonrobust than is the t test in general (though there are exceptions, and it must be kept in mind that the F statistics investigated were generally based on equal sized samples, while the majority of the investigated t statistics were not). Nevertheless, under nonnormality alone the approach of the FXX curves to their asymptotic value of 1.00 was very slow and "chance" crossings of this asymptote by the more stable curves (α = .05 or .01) occurred only at or above N = 128 (and in two of the three graphs only at or above N = 512). A rather mild degree of heterogeneity of variance (relative to what may easily be encountered in practice) spuriously inflated the probability of a

Type I error by an increment of roughly 10 to 300 percent, depending upon α, the number of samples drawn, and the relative variance of the population yielding the single sample, but virtually irrespective of sample sizes (see FYB curves). If a test deserves to be called robust under these circumstances, it clearly deserves also (and simultaneously) to be called a rough, approximate test. And Figures 55-70 clearly show that far greater degrees of nonrobustness than were found for F in the present study may be encountered when variances are more unequal, especially when a relatively small proportion of the samples are drawn from the population with the larger variance. In the behavioral sciences the experimenter, prior to collecting data, generally has little or no notion as to the degree of heterogeneity which may be present nor is he likely even to know with which treatments the larger variances will be associated. (Nor is he much better off after data have been collected unless the large amounts of data necessary for close estimation of population variances have been obtained.) One need only reflect upon these facts to realize that far greater hazards to the robustness of F lurk in the typical psychological experiment than have been explored in the present study.

VII Manner in which Individual Factors Influence Robustness

It has been shown that robustness is influenced by a multitude of factors whose individual effects may operate in directions and strengths which differ greatly among themselves and which, individually, may vary greatly with some other factor such as absolute sample size. Thus, robustness is the net effect of a combination of interacting factors. This raises the question of whether there is <u>any</u> statement about robustness in general which does not require particularistic qualification to make it true, i.e., any generalization which holds for all cases. Consider the statement, "The greater the extent or degree of the violation of an assumption (all other factors being held constant), the greater the detrimental effect upon robustness." Naively, one might suppose that if a safe general statement can be made about robustness, this is it. However, we have already seen that the robustness of the one-sample Z test appears to be more strongly affected by position on the symmetry-asymmetry dimension than by degree-of-fit-to-normality when the normality assumption is violated; and, specifically, the one-sample Z test is <u>far</u> more robust when samples are drawn from the extremely nonnormal, but, symmetrical, X-X population than when they are drawn from the far more nearly normal, but asymmetrical, X-Y population, and more robust than when they are drawn from the quasinormal, but slightly skewed, Y-A population. Nor are examples restricted to the normality assumption. "Increasing" heterogeneity of variance from a ratio of one to one to a ratio of four to one greatly increased the robustness of certain t tests (compare the right-tailed robustness of TXY 2N2N with the left-tailed robustness of TYA 2N2N and vice versa). Consider the statement, "Other things being equal, robustness increases with absolute sample size." This report is replete with contradictory examples, both for specified ranges of N (right tail of TXA N2N at $64 \leq N \leq 512$) and, apparently, in general (right tail of TXA N3N, right tail of RXY 2N2N). The statement, "Other things being equal, the smaller the one-tailed significance level the less robust the test." finds exceptions in cases where $\bar{\rho}$ passes from one side of α to the other (see right tail of TXA N2N at $N = 16$ where the .001 level is more robust than the .01 level, which is more robust than an α of .05; also note that ZX at $N \geq 16$ is more robust at $\alpha = .20$ than at smaller or larger αs). Statements that, other things being equal, robustness is greater for equal-sized than for unequal samples, for two-tailed than for one-tailed tests, for two-sample than for one-sample tests, for Z than for t, for F than for t, for nonnormality than for heterogeneity of variance, for identical than for mixed population shapes, or even for fewer violations of assumptions, all have exceptions (sometimes in abundance) which appear in the reported data of the present study. Thus, not only is it impossible to make unqualified generalizations about robustness which are invariably true, but it is apparently also impossible to make such statements about the qualitative effect upon robustness of any single one of the factors which contribute to it. This is equivalent to saying that none of the factors which influence robustness do so independently, i.e., produce effects which do not interact with the effects produced by other factors.

Clearly, therefore, it is quite futile to attempt to summarize the facts about robustness in a single pithy maxim, to be memorized and blindly followed by lay statisticians educated in the "cookbook" tradition. Nor will a handful of such maxims do the trick. To meet the needs of the "cookbook" statistician one would have to "catalogue" robustness under each of a huge number of combinations of sampling and testing conditions. However, it is possible, in certain cases, to show, how each factor exerts its influence upon robustness and therefore provide the information with which robustness effects can be partially anticipated on a logical basis:

Nonnormality. When the sampled population is nonnormal, the sample mean has a distribution which departs from its normal-theory distribution to a degree which appears to be highly correlated with the asymmetry of the population, and the extent of the departure diminishes (toward a perfect fit) with increasing sample size. The sample variance departs drastically from its normal-theory distribution if nonnormality is appreciable, and the extent of the departure does not diminish greatly with increasing sample size.

Heterogeneity of Population Shape. The robustness of a mean appears to be highly correlated with the degree of symmetry in the sampled population. The difference between two means, $\bar{X}_1-\bar{X}_2$, may be regarded as the mean of N values of a single variable, $V = (1/c_1)\sum_1^{c_1} X_1 - (1/c_2)\sum_1^{c_2} X_2$. If both sampled populations are symmetric (about the same point), the distributions of the means, $(1/c_1)\sum_1^{c_1} X_1$ and $(1/c_2)\sum_1^{c_2} X_2$, will each be symmetric (about the same point), insuring that the distribution of V will also be symmetric, irrespective of relative sample sizes. Also, if the two sampled populations are identical and sample sizes are equal, the distribution of $V = X_1-X_2$ will be symmetric, irrespective of whether or not the sampled populations are symmetric. Therefore heterogeneity of population shape should tend to be least detrimental to the robustness of $\bar{X}_1-\bar{X}_2$ when both populations are symmetric or when samples are of equal size and the two populations are identical

Heterogeneity of Variance. When population variances are unequal and samples have unequal sizes, the denominator of t is a biased estimate of the standard error of its numerator, and the denominator of the analysis-of-variance F test does not estimate the same quantity as its numerator. As a result of these biases in estimation, $\bar{p}$ approaches a limiting value different from α.

Equating sample sizes removes these biases, but the t and F distributions still do not have their normal-theory variances, except in the case of the t test when sample sizes are infinite. The variance of the F distribution tends to be inflated by heterogeneity of population variance when F is based upon more than two equal-sized samples. And when there are only two different populations, the distortion tends to be greatest when a small, but appreciable, minority of the samples come from the population with the larger variance. The result of these distorted variances is that $\bar{p}$ approaches a limiting value different from α. (The only practical case where $\bar{p}$ approaches α as a limiting value is when t, or $F = t^2$, is based upon two equal-sized samples.)

Other things, such as sample size, being equal, the extent to which a statistic is influenced by the shape of the distribution of one sample mean or variance rather than by that of another, is proportionate to the ratio of the respective population variances (or standard deviations). Therefore the robustness or nonrobustness of each sample's mean or variance is weighted (among other things) by its population's (relative) variance (or standard deviation) in the influence which it brings to bear upon the distribution of the test statistic.

Absolute Sample Size. As absolute sample size increases (a) the sample mean becomes more and more nearly normally distributed, and therefore robust, (b) the variance, σ^2/N, of the distribution of the sample mean diminishes and therefore so does the weight with which that mean influences the shape of the distribution of a two-sample Z or t statistic to which it contributes, provided that the size of the other sample remains constant, (c) the robustness of the sample variance is relatively unaffected (at least in the present series of studies) but the variance of the sample variance diminishes and consequently, (other things, such as population variances, being equal), the smaller one sample is relative to the others the more extreme its estimate of the population variance is likely to be and the less weight it receives in contributing to the pooled variance estimate in the denominator of t or F (but the more weight it receives in influencing the distribution of the two-sample F ratio between two estimates of population variance), (d) the smaller the variability of the denominator of t or F tends to become relative to the variability of the numerator and therefore the smaller the weight received by the shape of the denominator's distribution, which is highly nonrobust at all N's, relative to that received by the shape of the numerator's distribution, whose robustness improves with increasing N, in influencing the shape of the distribution of t or F.

Relative Sample Size. If N_1 increases while N_2 is held constant, (a) the robustness of $\bar{X}_1$ relative to that of $\bar{X}_2$ increases but its variance σ_1^2/N_1 decreases relative to the constant σ_2^2/N_2 and consequently the increased robustness of $\bar{X}_1$ exerts less influence than before upon the shape of the distribution of the Z or t statistic to which $\bar{X}_1$ and $\bar{X}_2$ both contribute, (b) the variance of $\hat{\sigma}_1^2 = \frac{\sum_1^{N_1}(X_1-\bar{X}_1)^2}{N_1-1}$ diminishes relative to that of $\hat{\sigma}_2^2 = \frac{\sum_1^{N_2}(X_2-\bar{X}_2)^2}{N_2-1}$ with the result that $\hat{\sigma}_1^2$ is likely to be less extreme, i.e., a more accurate estimate (due to its reduced variance), and receives more weight (since each sum of squares is weighted, in effect, by its degrees of freedom) in contributing to the pooled variance estimate in the denominator of t or F, but receives less weight, relative to that received by $\hat{\sigma}_2^2$, in influencing the shape of the distribution of $R = \hat{\sigma}_1^2/\hat{\sigma}_2^2$, since, its influence upon that statistic is "weighted" by its (reduced) variance but not by its (increased) degrees of freedom. If the sampled populations are identical but asymmetric, the variable $V = \left(1/c_1\right)\sum_1^{c_1} X_1 - \left(1/c_2\right)\sum_1^{c_2} X_2$ will be symmetrically distributed only (in general) when $N_1 = N_2$ (at which point V becomes simply the difference-score X_1-X_2). Therefore the mean $\bar{X}_1$-$\bar{X}_2$, of N values of V should tend to be maximally robust when $N_1 = N_2$ and, consequently, so should the two-sample Z test and the numerator of two-sample t. Also, if all samples have the same size the denominator of two-sample t remains an unbiassed estimate of the standard error of the numerator and the denominator and numerator of the analysis of variance F statistic estimate the same value even if variances are heterogeneous. Therefore equality of sample size tends to reduce the deleterious effects of violation of the assumption of homogeneity of variance for the t and F tests. If population variances and sample sizes are both unequal, the denominator of two-sample t is a biassed estimate of the standard error of the numerator, and the distribution of the F statistic does not even have its normal-theory location; the greater the proportionate discrepancy between sample sizes, the greater tends to be the extent of the bias in t.

Number of Samples. Since one-sample tests draw a single sample from a single population, there can be no heterogeneity of population shape or variance and no relative sample size and such factors therefore cannot interact with violation of the test's assumptions. Furthermore there tend to be fewer assumptions to violate. The robustness of one-sample tests therefore tends to be less complicated and easier to predict than is the case for multi-sample tests. In a sense these advantages are shared by the two-sample Z test since that test can generally be reduced to the statistical equivalent of a one-sample Z test, (especially when one sample size is an integral multiple of the other). Multi-sample tests however enjoy a number of advantages. Asymmetry in the sampled population appears to be one of the worst types of violation of the normality assumption so far as robustness is concerned. Nothing can be done about it for one-sample tests. However, for two-sample tests sampling from identical asymmetric populations, the deleterious effects of asymmetry can be removed by taking samples of equal size. This makes the two-sample Z or t test equivalent to a one-sample test based on observations drawn from a population of (uncorrelated) difference-scores, and differences between observations drawn from identical populations are symmetrically distributed. Furthermore when the sampled population is nonnormal the distribution of the sample variance is quite nonrobust, the variance tending to assume both very large and very small values at far greater than normal-theory frequency. The denominator of the one-sample t test contains a single sample variance while that of the two-sample t test contains a sort of weighted average of two sample variances and that of the F test contains a sort of weighted average of as many variances as there are samples. Thus, the one-sample t test will have an extreme denominator just as often as there is an extreme sample variance, but in the case of the two-sample t test or of the F test the effect of an extreme sample variance may be counteracted by that of a variance of the opposite extreme or moderated by that of a more central variance. As a result, other things being equal, the denominator of multi-sample F should be more robust than that of two-sample t which should be more robust than that of the one-sample t test.

Type of Test. Since the sample mean is more robust than the sample variance, tests using only sample means have an advantage over those using both sample means and sample variances, and the latter are likely to prove more robust, under otherwise comparable conditions, than those using only sample variances, i.e., variances of original observations.

Significance Level. When assumptions are violated the fit between the true and normal-theory distributions of the test statistic tends to become increasingly worse at increasingly remote tail regions. Therefore the smaller, i.e., the more extreme, the nominal significance level (for a test involving specified tails) the greater tends to be the relative departure of the true significance level from it, and the more nonrobust the test tends to be. (Exceptions tend to be rare and highly idiosyncratic.)

Location of Rejection Region. The normal-theory distributions of Z and t are symmetrical. However, if a sampled population is asymmetrical (and samples are of unequal size) the true distributions of Z and t will also be asymmetrical. As a result the fit between the true and normal-theory distributions may be quite different at symmetrically opposite points of the normal-theory distribution. The true significance level, $\bar{\rho}$, corresponding to a normal-theory significance level of α may be quite close to α for a left-tailed test at nominal significance level of α, but far removed from α for a right-tailed test at the same nominal significance level. If $\bar{\rho}$ deviates from α in the same direction at the two tails, then the robustness of a two-tailed test at nominal significance level of 2α will be a sort of average of the robustness of the left-tailed and right-tailed tests at one-tailed significance levels of α. However, if $\bar{\rho}$ deviates from α in opposite directions at the two tails, the robustness of a two-tailed test at nominal significance level of 2α may be better than that for either one of the one-tailed tests at the α level. Thus population asymmetry makes it possible for two-tailed tests to be far more robust than one-tailed tests at half their significance level (and, in fact, if a two-tailed test at level 2α is far more robust than a one-tailed test at level α, the superiority is extremely likely to persist when the one-tailed test is at level 2α). In any case, the two-tailed test at level 2α will be more robust than the less robust of the two one-tailed tests at level α and this will be the case whether $\bar{\rho}$ deviates from α in the same or in opposite directions at the two tails. If the sampled populations are symmetrical the true distributions of the Z and t statistics will also be symmetrical. Therefore the deviation of $\bar{\rho}$ from α at the right tail will be exactly the same, both in direction and extent, as at the left tail. The robustness of a two-tailed test at level 2α is therefore identical to that of either one-tailed test at level α. And since robustness tends to diminish with decreasing α's, the robustness of the two-tailed test at level α should be less than that of either of the one-tailed tests at the same level, α. Consequently when the sampled populations are symmetrical the two-tailed Z and t tests tend to be less robust than one-tailed Z and t tests performed at the same significance level.

VIII General Comments

As pointed out in the earlier study, the question of the relative sensitivity of a test to violation of its various assumptions is fairly meaningless unless one is willing to specify exactly "how much" violation and under exactly what sampling conditions (i.e., what sample sizes, what significance levels, what rejection regions, etc.). The robustness of the test depends upon the specific situation. And, since in the present study a single nonnormal population shape was investigated for the t, R, and F tests, the results might be attributed to the inadvertent investigation of a population whose shape was uniquely conducive to nonrobustness. However, such a conclusion is impugned by the results of the one-sample Z tests based on observations drawn from one of six different population shapes. The shapes varied greatly both in degree of symmetry and in fit to normality. However only in the case of the exactly symmetrical X-X population did results for Z approach a really conservative criterion of general robustness at an early value of N. Therefore since (a) exactly symmetrical populations of absolute scores are rare if not nonexistent in the behavioral sciences, and (b) the one-sample Z test tends to be considerably more robust than other one-sample statistics (such as the analogous t-test) under corresponding conditions of sampling, it seems reasonable to conclude that had a different, asymmetrical population shape (i.e., a shape likely to be encountered for absolute scores in the behavioral sciences) been substituted for that of the X population, results would still have been quite damaging to the concept of general robustness.

The term "robust", as generally used, is relatively meaningless because (a) there is no accepted definition of, or criterion for, robustness (e.g., there is no consensus regarding the deviation of $\bar{p}$ from α which justifies the use of the term), (b) even if there were an accepted criterion of robustness, the term, when applied to a particular test, requires qualification specifying the conditions under which the criterion is met, and such qualification rarely accompanies the use of the term. In addition to the need for explicitness along these lines, results of the present study suggest the need for certain new qualifying terms, which might greatly reduce ambiguity. A given test, performed under given and stated violations of assumptions, might be called "ultimately completely robust", (as has been done in the present study) if when its smallest sample reaches infinite size the test statistic has exactly its normal-theory distribution. By specifying complete robustness at infinite sample size, it obviates specification of either absolute sample sizes, the value of α, or the location of the rejection region. Another term needed is "monotonically increasingly robust". For a test which is monotonically increasingly robust, $\bar{\rho}$ approaches α monotonically (but α does not have to be the limiting value of $\bar{\rho}$) as the absolute size of the samples increases while their relative sizes remain constant. Finally, for a test which is "bilaterally nonrobust" $\bar{\rho}$ exceeds α under some conditions of sample size and is exceeded by it under others. In the opinion of the writer the absence of ultimate complete robustness or of monotonically increasing robustness (especially when the deviation from the latter takes the form of bilateral nonrobustness) tend, when they occur, to render robustness an unmanageably complex concept unless a firm upper bound can be placed upon the value of $|\bar{\rho}-\alpha|$. Otherwise stated, in order to make good use of "robustness" the experimenter, at the very least, needs to have either the qualitative information that robustness constantly increases toward ultimate, complete robustness as N increases, (i.e., that $\bar{\rho}$ deviates from α in a single consistent direction, monotonically approaching α as N increases), or the quantitative information that $\bar{\rho}$ will never deviate from α by more than a specified amount which falls within his subjective tolerance region. Since the latter information is seldom available (except at N = infinity!) for the specific case in which he is interested, and since, when it is available, it is likely to require much more precise information about the shapes and variances of his populations than he is likely to possess, he generally will require the former. But, empirical data in the present study showed thatat low N's and α's of .20 or .30, ZX was not monotonically increasingly robust but was, in fact, bilaterally nonrobust.

And the one-sample Z test is the least complicated statistic investigated, so far as robustness is concerned! Likewise there were many instances where the two-sample t test was not monotonically increasingly robust or was bilaterally nonrobust, and these instances involved the testing tails and occurred when variances were homogeneous as well as when they were unequal. And a similar statement could be made about F at low N values. Perhaps the Z and F statistics cease to be bilaterally nonrobust above some fairly small N value. But, if so, at what N value? Thus the concept of robustness, if properly used, is plagued by particularistic considerations. Unfortunately those applied statisticians who are most enamoured of the term appear to be those who are least aware of its particularistic nature.

SUMMARY

Many parametric test statistics are said to be robust against violation of certain (or all) of their assumptions. By this is meant that the true significance level of the test when the assumption is violated is a sufficiently close approximation to the nominal significance level, i.e., the level which would be the true significance level if all assumptions were met. Let α be the nominal, i.e., the normal-theory, significance level of the test, and let S_α be that value of the statistic whose true cumulative probability in the normal-theory distribution of the statistic (i.e., when the null hypothesis and all assumptions are true) is α. Thus S_α is the boundary of the rejection region and is the value of S listed under a significance level of α in the conventional tables for the test statistic. Regardless of whether assumptions are violated or not, the test will reject whenever the experimentally obtained value of S falls on the "rejection region side" of S_α. But when assumptions are violated the true cumulative probability (i.e., the true significance level) of S_α will no longer be α. Let $\bar{\rho}$ be the cumulative probability of S_α in the <u>true</u> distribution of the statistic when the assumption is violated, and let ρ be the chance-influenced cumulative probability of S_α in an <u>empirical</u> sampling distribution of S's, under exactly the same circumstances of assumption-violation, sampling, etc. Then $\bar{\rho}$ is the true significance level, and ρ is an estimate of it, when the test is performed, in violation of the assumption, at an alleged significance level of α. Finally, let $\bar{\rho} - \alpha = \bar{\epsilon}$ and $\rho - \alpha = \epsilon$, so that ϵ is an estimate of $\bar{\epsilon}$. The robustness of a test, therefore, depends on the value of $\bar{\epsilon}$. If $\bar{\rho}$ is a sufficiently close approximation to α, $\bar{\epsilon}$ will be "sufficiently" small. Also $\bar{\epsilon}/\alpha$ will be "sufficiently" close to zero and $\bar{\rho}/\alpha$ (or $\alpha/\bar{\rho}$) will be "sufficiently" close to 1.00, and these latter have the advantage of expressing deviations from α in units of α (or their reciprocal).

But there is no generally accepted criterion distinguishing quantitatively between robustness and nonrobustness. Not only is there no consensus as to how small $\bar{\epsilon}$ or $\bar{\epsilon}/\alpha$ should be or how close $\bar{\rho}/\alpha$ or $\alpha/\bar{\rho}$ should be to 1.00, but statements claiming robustness for a given test seldom include (or even imply) any objective definition or quantitative criterion for robustness whatever. The statements of such claims are even more remiss. They often appear in the form "The blank test is robust against violation of the dash assumption". Such an allegation represents pure semantic chaos. Robustness is not defined, nor is the extent of the violation. However the discrepancy between $\bar{\rho}$ and α depends upon the extent of the violation. It also depends upon a host of other factors not even mentioned in the allegation of robustness, such as the value of α, the location of the rejection region, the number of samples used by the test, their absolute and relative sizes, the relative shapes and variances of the sampled populations, and interactions between the factors named, between the violations of assumptions and between factors and violations. Statements of the type quoted, therefore, are worse than worthless. They have, however, given rise to a belief among the statistical laity that most, if not all, parametric tests are astoundingly impervious to most, if not all, violations of assumptions irrespective of the prevailing sampling or testing conditions, except perhaps that samples should not be extremely small. A previous study exposed these fallacies for the one-sample Z, t and χ^2 (variance) tests. The present study gathered additional data relevant to the one-sample Z and t tests and extended the exposure to the two-sample Z and t tests, the F ratio of two sample variances, and the analysis-of-variance F test.

In the present study there were four original populations, each of which was discrete and finite, had a mean, μ, of 100, and consisted of 100,000 units. The four populations are identified by the following letters

X   A very highly skewed population with mean μ and variance σ^2 and consisting only of integers.

Y As nearly normal a population as could be constructed with 100,000 integers with mean μ and variance σ^2, therefore differing from the nonnormal X population primarily in that it was essentially normal in shape.

A A population obtained by moving every variate-value of the X population halfway to the mean, μ. The A population is therefore identical to the X population except that its variance is $\sigma^2/4$ and its values can be midpoints between integers as well as integers. Thus the A population has the same shape as the X population but only one-fourth the variance.

B A population obtained by moving every variate-value of the Y population halfway to the mean, μ. The B population is therefore identical in shape (discrete quasi-normal) to the Y population but has only one-fourth the variance and consists of half-integral as well as integral values.

The X and Y populations are identical to those investigated in the previous study and designated there as X and Y respectively.

The one-sample statistics investigated in the present study are identified by a code of the form SP
where the first letter, S, stands for the type of statistic investigated of which there were two as follows:

Z one-sample Z statistic

T one-sample t statistic

The second letter stands for the population, X or Y, from which the sample was drawn.

The two-sample statistics investigated in this study are identified by a code of the form SPQ C_1NC_2N
where the first letter, S, stands for the type of statistic of which there were three as follows:

Z two-sample Z statistic

T two-sample t statistic

R two-sample F ratio between two sample estimates of population variance.

The second and third letters P and Q represent the populations, X, Y, A or B, from which the first and second sample, respectively, were drawn. C_1N and C_2N are the respective sizes of the first and second samples (i.e., C_1N observations were drawn from population P, and C_2N from population Q). C_1 and C_2 are integers which assume values of 1, 2, or 3. When one of the C's equals 1, the 1 is omitted in front of the N.

The three-sample and four-sample statistics investigated are identified by the code

$$\text{FPQ } C_1N, C_2NC_3NC_4N$$

where F stands for the analysis-of-variance F test for equal population means, C_1N, C_2N, C_3N, and C_4N are the sizes of the samples and P and Q are the populations contributing the samples whose sizes are listed to the left and to the right of the comma respectively. For three-sample tests, $C_4N = 0$ and C_4N is omitted from the code designation. When a C equals 1, the 1 is omitted. When P and Q are the same population, the comma is omitted.

TD 2N is a special case, standing for the two-sample t statistic based on matched pairs of observations. The test is based on 2N observations from the X population and 2N observations drawn from the Y population but correlated with the X observations by using the same random numbers to identify them.

(The codes given above are those used in the body of the report. A slight variation of these codes is used in the Appendix.)

For each of the following values of N, 2, 4, 8, 16, 32, 64, 128, 256, 512, and 1024, ten thousand sets of 4N pseudo-random numbers were generated by an IBM 7090 computer. Each such set of 4N pseudo-random numbers was used to identify the observations required to calculate a <u>single</u> value of <u>each</u> of the multi-sample statistics investigated in this study and identified by a unique set of code symbols. Thus for each of the 102 multi-sample statistics identified by a unique set of code symbols in this study (i.e., not simply for each general type of multi-sample statistic but for the unique set of sampling conditions under which it was obtained in specific cases) an empirical sampling distribution of 10,000 values was obtained at each of the ten values of N. (For the one-sample statistics empirical sampling distributions of 40,000 values from the present study were combined with those based on 10,000 values in the previous study to yield highly reliable sampling distributions of 50,000 empirically obtained values at each value of N $\leq$ 1024; sampling distributions based upon 10,000 values each were also obtained in the present study for the one-sample statistics at N's of 2048 and 4096.) For a given statistic, with unique code symbol, the ten sampling distributions corresponding to the ten different values of N (which are $\leq$ 1024) are completely independent. However, for a given value of N the empirical sampling distributions of the various statistics are not independent, and, in fact, their dependence tends to be maximized by the method used to identify observations for the component samples contributing to a single value of a statistic. The empirical sampling distributions of the various statistics at a given value of N, therefore, tend to be maximally comparable, owing to maximal similarity between the causes of their chance deviations.

When all samples are drawn from the Y population, statistics based upon them should have very nearly their normal-theory distributions, except perhaps at the smallest values of N (at which values the discreteness of the Y population sometimes appreciably affects the sampling distributions) provided sampling has been properly random. Thus, such statistics provide "control" distributions. And for the ZY statistic, the sampling distribution should be symmetrical at <u>all</u> values of N, provided that sampling has been sufficiently random. These facts were exploited to test (indirectly) the randomness of sampling at every value of N. Results of such tests indicated quite satisfactory randomness of sampling at every value of N.

When samples are of equal size, the two-sample Z statistic reduces mathematically to a one-sample Z statistic based upon a single sample of difference-scores. And the latter is simply a linear transformation upon the mean of the sample of difference-scores. The sampling distribution of this mean (or of any linear transformation upon it) at various increasing sample sizes shows the Central Limit Effect for means of samples of those sizes drawn from that population of difference-scores. Thus the empirical sampling distributions of ZXX 2N2N, ZXY 2N2N, ZXA 2N2N, ZXB 2N2N, and ZYA 2N2N show the Central Limit Effect upon the means of 2N observations drawn from the X-X, X-Y, X-A, X-B, and Y-A, difference-score populations respectively. By pairing every value in one absolute-score population with every value in a second such population, obtaining the difference between the two values and the product of the point probabilities of the two values, and summing these products over all pairs having the same difference-score, one obtains the distribution of the population of possible difference-scores. In this manner, the shapes of the X-X, X-Y, X-A, X-B, and Y-A difference-score populations were determined and graphed. Results of this study therefore showed the Central Limit Effect upon the means of samples from each of six populations, X, X-X, X-Y, X-A, X-B, and Y-A, differing greatly in general shape, in degree of symmetry, and in fit to a normal distribution with the same mean, variance, and area. It was found that the nonrobustness of the sample mean (i.e., the extent to which the Central Limit Effect had failed to produce normality) was much more strongly related to the asymmetry than to the nonnormality of the sampled population. There was an almost perfect linear relationship between the coefficient of skewness, μ_3/σ^3, for the sampled population and the nonrobustness of the sample mean (as measured by $|\frac{\rho-\alpha}{\alpha}|$ at sample size of 16, averaged over left and right tailed α's of .05 and .01). The nonrobustness of the sample mean was also impressively related to the following indices of asymmetry: (a) the distance in sigma units between interpolated median and mean of the sampled population, (b) the sigma-unit distance between mode and mean, (c) the proportion of the total population variance contributed by units lying in asymmetric area, i.e., areas having no mirror image on the other side of the axis through the population mean. In cases (a) and (c) the rank order correlation was perfect, and in all cases the relationship seemed very nearly linear. The relationship between nonrobustness of the sample mean and nonnormality (when the latter is measured by the proportion of a population's area which does not lie within a normal distribution having the same mean, variance, and area) was feeble and lackluster by comparison, and what relationship there was seems likely to be attributable to a correlation between measures of nonnormality and of asymmetry. A clear test of whether asymmetry or nonnormality is the critical factor was provided by the X-X population which is the least conducive to nonrobustness, the least asymmetric (in fact, it is exactly symmetric), but among the most nonnormal.

Results were presented in great detail. For every statistic and every combination of sampling conditions investigated, tables in the Appendix list the ρ's corresponding to left-tailed, and to right-tailed, α's of .0005, .001, .005, .01, .025, .05, .10, .20, .30, .40, and .50, and graphs in the Results section show the ratio between ρ and α as a function of N for α's of .05, .01, and .001. A single aspect of these results is summarized in Tables V and VI. Table V gives the smallest N value at which ρ finally entered and remained within a tolerance region of $\alpha \pm \alpha/2$. Based on a sampling distribution of 10,000 samples, the .95, .99, and .999 tolerance regions about an α of .05 are $\pm .086\ \alpha$, $\pm .112\ \alpha$, and $\pm .144\ \alpha$, respectively; about an α of .01 they are $\pm .200\ \alpha$, $\pm .260\ \alpha$, and $\pm .330\ \alpha$, respectively; and about an α of .001 they are $\pm .600\ \alpha$, $\pm .800\ \alpha$, and $\pm 1.000\ \alpha$, respectively. Therefore when $\bar{\rho}$ is nearly equal to α, ρ is unlikely to depart from the tolerance region $\alpha \pm \alpha/2$ if $\alpha = .05$ or $\alpha = .01$, but not if $\alpha = .001$. For this reason Table V is limited to α's of .05 and .01. It is emphasized that the criterion does not assume either that the "true" value or that the limiting value of $\bar{\rho}$ is α. Nor is it implied that a test is robust when $.5\ \alpha < \rho < 1.5\ \alpha$. In fact a departure of ρ from α by (1/2) α was chosen as the criterion upon which to construct the table because this was the most liberal criterion of robustness the writer could bring himself to take seriously. (Note that ρ cannot be less than zero and therefore cannot deviate below α by more than 100% of α). Three characteristics of Table V should be noted. First, statistics based upon samples from the same pair of populations are arranged in order of increasing size of sample drawn from the first-listed population relative to the size of that drawn from the second. Second, statistics occupying the same row are based upon the same total number of observations, drawn in the same proportion from the members of the same pair of populations. Thus, corresponding entries on the same row permit comparison of the Z statistic with the T statistic under identical sampling and test conditions or comparison of either Z, T, or F under maximally similar sampling and test conditions (provided that a right-tailed F test is understood to "correspond" to a two-tailed Z or T test). Also, since the two-tailed T test is equivalent to a right-tailed, two-sample F test, comparison of the two-tailed T entry with the corresponding F entry on the same row shows how robustness of the F test varies when the total number of observations (as well as the number of observations drawn from each population) is constant but the number of samples among which the constant number of observations from a single population are divided varies. Finally, "N" is not necessarily a sample size, and certain allowances may have to be made in order to make certain statistics comparable. For example ZX is based on one sample of N observations, while ZXX 2N2N is based upon two samples

TABLE V

SMALLEST N VALUE COMMENCING WITH WHICH $\hat{p}$ ALWAYS DEVIATED FROM α BY LESS THAN 50% OF α (SO THAT $.5\alpha < \hat{p} < 1.5\alpha$)

Statistic	α = .05 Left Tail	α = .05 Two Tail	α = .05 Right Tail	α = .01 Left Tail	α = .01 Two Tail	α = .01 Right Tail	Statistic	α = .05 Left Tail	α = .05 Two Tail	α = .05 Right Tail	α = .01 Left Tail	α = .01 Two Tail	α = .01 Right Tail	Statistic	α = .05 Right Tail	α = .01 Right Tail
							Nonnormality Alone									
ZX	32	2	4	256	32	256	TX	256	128	128	2048	512	1024			
ZXX N3N	4	2	16	128	16	64	TXX N3N	16	2	32	128	2	128	FXX NNNN	8	8
N2N	4	2	2	32	16	2	N2N	2	2	16	32	8	128	NNN	16	16
2N2N	2	2	2	4	4	4	2N2N	4	8	4	16	32	16	2NNN	2	8
2NN	4	2	2	32	16	2	2NN	2	2	16	32	8	128	NNNN	16	16
3NN	4	2	16	128	16	64	3NN	16	2	32	128	2	128	NNNNN	8	8
							Nonnormality and Mixed Shapes									
ZXY N3N	16	2	4	128	16	128	TXY N3N	2	4	2	16	8	8	FXY N, NNN	2	8
N2N	16	2	4	128	16	128	N2N	2	2	8	64	2	64	N, NN	4	16
2N2N	2	2	2	32	2	16	2N2N	16	4	8	128	32	64			
2NN	2	2	2	2	2	4	2NN	32	16	2	256	64	128	NN, N	4	32
3NN	2	2	2	2	2	2	3NN	32	16	2	128	32	32	NNN, N	4	16
							Heterogeneity of Variance Alone									
							TYB N3N	—	—	—	—	—	—	FYB N, NNN	2	—
							N2N	—	—	—	—	—	—	N, NN	2	—
							2N2N	2	2	2	2	2	2			
							2NN	—	—	—	—	—	—	NN, N	2	8
							3NN	—	—	—	—	—	—	NNN, N	2	16
							Nonnormality and Heterogeneity									
ZXA N3N	32	2	4	256	16	128	TXA N3N	—	—	—	—	—	—	FXA N, NNN	64	—
N2N	32	2	4	256	32	256	N2N	—	—	—	—	—	—	N, NN	64	—
2N2N	8	2	2	32	8	32	2N2N	64	32	32	256	64	256			
2NN	2	2	2	2	8	16	2NN	32	—	—	—	—	—	NN, N	16	32
3NN	2	2	2	4	8	4	3NN	—	—	—	—	—	—	NNN, N	2	16
							Nonnormality, Heterogeneity, and Mixed Shapes									
ZXB N3N	32	2	4	256	32	128	TXB N3N	—	—	—	—	—	—	FXB N, NNN	128	—
N2N	32	2	4	256	16	128	N2N	—	—	—	—	—	—	N, NN	128	—
2N2N	8	2	2	64	8	32	2N2N	64	32	64	512	128	256			
2NN	4	2	2	64	4	32	2NN	32	—	—	—	—	—	NN, N	16	64
3NN	2	2	2	8	4	16	3NN	—	—	—	—	—	—	NNN, N	8	32
ZYA N3N	2	2	2	2	2	2	TYA N3N	—	—	—	—	—	—	FYA N, NNN	8	—
N2N	2	2	2	2	2	2	N2N	—	—	—	—	—	—	N, NN	8	—
2N2N	2	2	2	2	2	2	2N2N	2	2	4	2	8	8			
2NN	2	2	2	16	4	4	2NN	—	—	—	—	—	—	NN, N	4	16
3NN	2	2	2	32	4	32	3NN	—	—	—	—	—	—	NNN, N	2	16
							Nonnormality Alone									
RX 2N2N	—	—	—	—	—	—										
							Nonnormality and Mixed Shapes									
RXY 2N2N	—	—	—	—	—	—	TD 2N	4	2	2	32	8	2			

TABLE VI

TERMINAL VALUE, AT N = 1024, OF $\hat{p} - \alpha$ EXPRESSED AS A PERCENTAGE OF α

(FOR THOSE STATISTICS FOR WHICH $\hat{p}$ DOES NOT APPROACH A LIMITING VALUE OF α)

Statistic	α = .05 Left Tail	α = .05 Two Tail	α = .05 Right Tail	α = .01 Left Tail	α = .01 Two Tail	α = .01 Right Tail	α = .001 Left Tail	α = .001 Two Tail	α = .001 Right Tail	Statistic	α = .05 Right Tail	α = .01 Right Tail	α = .001 Right Tail
					Heterogeneity of Variance Alone								
TYB N3N	129	208	129	340	481	340	1020	1386	1020	FYB N, NNN	26	100	300
N2N	77	114	77	100	224	100	408	620	c.3	N, NN	16	88	260
2NN	-57	-49	-57	-75	-77	-75	-88	-88	-88	NN, N	8	36	110
3NN	-75	-88	-75	-88	-90	-90	-98	-100	-98	NNN, N	7	8	90
					Heterogeneity and Nonnormality								
TXA N3N	141	188	114	365	474	303	1330	1560	800	FXA N, NNN	30	121	310
N2N	91	118	64	226	347	150	710	720	290	N, NN	17	73	260
2NN	-46	-64	-42	-62	-75	-52	-90	-100	-88	NN, N	9	33	80
3NN	-66	-63	-70	-80	-86	-86	-100	-100	-100	NNN, N	12	31	80
					Heterogeneity, Nonnormality, and Mixed Shapes								
TXB N3N	130	188	111	377	488	280	1010	1740	780	FXB N, NNN	33	122	320
N2N	88	114	66	260	272	126	770	680	380	N, NN	19	80	270
2NN	-44	-68	-61	-80	-71	-80	-88	-88	-100	NN, N	8	37	110
3NN	-68	-82	-80	-86	-91	-90	-90	-90	-100	NNN, N	10	33	80
TYA N3N	127	208	134	330	478	327	1080	1340	1040	FYA N, NNN	30	118	380
N2N	76	116	88	174	333	164	480	870	800	N, NN	15	70	280
2NN	-67	-66	-66	-80	-82	-74	-90	-100	-90	NN, N	8	30	100
3NN	-76	-86	-75	-92	-90	-90	-100	-100	-100	NNN, N	6	16	90
					Nonnormality Alone								
RXX 2N2N	400	773	417	1003	2082	1094	10480	13840	10080				
					Nonnormality and Mixed Shapes								
RXY 2N2N	300	515	286	1062	1680	807	6040	8080	4080				

Note: An entry of -100 represents the maximum possible deviation of $\hat{p}$ below α and is an absolute lower bound for table entries.

BLANK PAGE

each consisting of 2N observations. Thus each sample of ZXX 2N2N is twice the size of the single sample contributing to ZX, and there are four times as many observations. Also, the maximum value N can assume is 4096 for the one-sample ZX and TX statistics and 1024 in all other cases. For some of the statistics listed in Table V, ρ does not approach a limiting value of α. This is true of the T test when variances are heterogeneous and samples are of unequal size and of the multi-sample F test when variances are heterogeneous. Table VI lists, for these statistics, the value of ρ-α as a percentage of α at N = 1024, the largest N value investigated for these tests. It is not implied that ρ at N = 1024 is a good estimate of $\bar{\rho}$ at N = ∞, nor is it implied that $\bar{\rho}$ at N = ∞ is necessarily closer to α than is $\bar{\rho}$ at N = 1024. Table VI implies only that, for the statistics listed, ρ is a biassed estimate of α even at infinite sample sizes and that the extent of the bias is appreciable in all cases and great in most at an extremely large finite sample size.

Consider now the individual factors contributing to the robustness of the various tests. The general robustness of a statistic may be defined as the overall goodness of fit between the true distribution of the statistic (under assumption-violating conditions) and the corresponding normal-theory distribution. Its specific robustness may be defined as the goodness of fit between $\bar{\rho}$ and α and represented by an index such as $\bar{\rho}-\alpha$, $(\bar{\rho}-\alpha)/\alpha$, $\bar{\rho}/\alpha$ or $\alpha/\bar{\rho}$. Thus while general robustness is influenced only by factors associated with sampling, specific robustness is, in addition, influenced by factors associated with testing, i.e., influenced by factors associated with α, namely the size of the significance level and the location of the rejection region. It is an empirical fact that, for all statistics investigated, goodness of fit between empirical and normal-theory distributions tended to be poorest at the testing tails and to worsen as increasingly remote tail regions are reached. (Exceptions were rare and highly transitory.) Thus $|(\rho-\alpha)/\alpha|$ tended, when all other factors were held constant, to increase with decreasing tail values of α. If the sampled population for a one-sample Z or t test is symmetric, or if the sampled populations for a two-sample Z or t test are identical and sample sizes are equal, or if the sampled populations for a two-sample Z or t test are both symmetric and have a common axis of symmetry, the distribution of the test statistic will be symmetric, and a left-tailed test will be exactly as robust as a right-tailed test at the same nominal significance level, α. Furthermore, a two-tailed test at level α will be exactly as robust as a one-tailed test at level $\alpha/2$, and since robustness tends to diminish with decreasing α a two-tailed test will therefore tend to be less robust than a one-tailed test conducted at the same nominal significance level. If the sampled population for a one-sample Z or t test is asymmetric, or if the populations for a two-sample Z or t test are not coaxially symmetric or if they are not identical when sample sizes are equal, the true distribution of Z or t will be asymmetric and $\bar{\rho}-\alpha$ will not necessarily have the same value at the two opposite tails. In such cases a left-tailed test may be far more or far less robust than a right-tailed test conducted at the same α level. If ρ deviates from α in the same direction at the two tails, the robustness of a two-tailed test at level 2α will be intermediate between the robustnesses of left-tailed and right-tailed tests at level α. However, if ρ deviates from α in opposite directions at the two tails, the one-tailed departures of ρ from α tend to cancel and a two-tailed test at level 2α may be far more robust than either a left or right tailed test at level α. Thus, robustness is influenced not only by the size of the significance level, but also by the location of the rejection region, i.e., whether left-tailed, right-tailed, or two-tailed. Such robustness-influencing factors are associated with testing, rather than sampling, and therefore are common to all of the tests investigated (except that the analysis-of-variance F test uses only one location of rejection region).

Also common to all of the tests investigated (if TD be regarded as a one-sample test) are the assumptions of random and independent sampling. These assumptions are unique in that (in a sense, at least) they are concerned with procedures rather than populations and can be met, at the option of the experimenter, by rigidly following certain rules of sampling. In this regard they differ from "population" assumptions, such as normality and homogeneity, over which the experimenter seldom has any direct control. The present report is primarily concerned with violation of the latter type of assumption and has not investigated violation of the former. Therefore the assumptions of random and independent sampling will be taken for granted as having been met in the discussion to follow, and interactions between their violations and other violations or factors will be ignored. Actually sampling is seldom completely random and often is not entirely independent and robustness may be greatly affected by violation of either of these two assumptions. Furthermore, effects of their violations can be large in their own right and can be expected to interact with those of other violations as well as with extraneous factors. Clearly, therefore, ignoring these assumptions biasses the discussion in favor of the conclusion that the tests are robust.

Turning now to the sampling factors influencing the robustness of the various tests, the one-sample Z statistic, $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$, is perhaps the least complicated statistic so far as robustness is concerned. Its formula contains a single variable, $\bar{X}$, and the test's only assumption about the sampled population is that it is normally distributed. If this assumption is true, then $\bar{X}$ will have a normal distribution with mean μ and variance σ^2/N, i.e., will have the distribution upon which the test is based, and in fact the "assumption" that the sampled population is normal may be validly replaced by the "assumption" that $\bar{X}$ is normally distributed with mean μ and variance σ^2/N. But, even if the original assumption that the sampled population is normal is false, the "substitute" assumption will be more and more nearly met as sample size increases, owing to the Central Limit effect. Thus there is a single assumption (other than independent random sampling) whose violation is complicated by a single factor associated with sampling, namely sample size. Consequently, in accordance with the Central Limit Theorem the general (although not necessarily the specific) robustness of one-sample Z should increase monotonically as sample size increases, and the limiting value of $\bar{\rho}$ is α so that Z becomes an exact test when sample size is infinite. Thus, the one-sample Z test makes the single, "population" assumption of normality, but, when that assumption is violated, its robustness is influenced by at least the following factors:

(a) "degree" of nonnormality (especially the asymmetric aspect) of the sampled population

(b) size, N, of sample

(c) size of significance level, α

(d) location of rejection region.

The two-sample Z test, $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$, assumes (with regard to populations) only that both sampled populations are normally distributed, and this assumption can be replaced by the functionally equivalent assumption that the true distributions of $\bar{X}_1$ and $\bar{X}_2$ are normal with means of μ_1 and μ_2 and variances of $\frac{\sigma_1^2}{N_1}$ and $\frac{\sigma_2^2}{N_2}$, respectively, since if this is the case Z will have its normal-theory distribution. Again, the Central Limit Theorem assures that even if the original assumption is false, the substitute assumption will be more and more nearly met as either or both sample sizes increase. Thus absolute sample size is again an influential sampling factor. However, this time there are complications. The distribution of Z is simply the pattern of its variability, and the variance of Z is a linear function of the variance of $(\bar{X}_1 - \bar{X}_2)$, which, in turn, is the sum of the variances of $\bar{X}_1$ and $\bar{X}_2$. Therefore the greater the variance of $\bar{X}_1$, say, relative to that of $\bar{X}_2$, the greater will the shape of the Z distribution tend to be influenced by the shape of the $\bar{X}_1$ distribution rather than by the shape of $\bar{X}_2$'s distribution. Thus, in a sense, any nonnormality (or normality) in the distribution of an $\bar{X}$ is weighted by the variance (or, perhaps more accurately, by the standard deviation) of that $\bar{X}$ in its influence upon the shape of the Z distribution. Since the variance of an $\bar{X}$ is σ^2/N, the influence of the $\bar{X}_1$ distribution's shape upon the Z distribution's shape tends to be greater the greater σ_1^2 is relative to σ_2^2 and the smaller N_1 is relative to N_2. Thus relative population variance and relative sample size enter as complicating factors when the normality assumption is violated. If the original populations are normally distributed with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, then $\bar{X}_1 - \bar{X}_2$ will be normally distributed with variance

$$\sigma^2_{\bar{X}_1-\bar{X}_2} = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2},$$

and this is all that the Z test actually requires. Therefore the assumption of normally distributed populations may be replaced by the substitute assumption that $\bar{X}_1 - \bar{X}_2$ is normally distributed. If both sampled populations are symmetric about the same axis, or if the sampled populations are asymmetric but identical in shape and variance and samples are of equal size, the distribution of $\bar{X}_1 - \bar{X}_2$ will be symmetric and, as a result, Z will tend to be robust. Hence relative population shape is yet another factor influencing the robustness of Z. Thus, even for the relatively simple two-sample Z test the situation with regard to robustness may be quite complex. Robustness is a function of a number of interacting factors and consequently prediction is hazardous. For example, as N_1 is increased while N_2 is held constant, does Z become more robust because of the increased robustness of $\bar{X}_1$, or less robust because of the diminished weight which $\bar{X}_1$ receives relative to $\bar{X}_2$ in influencing the distribution of Z? If one sample size is an integral multiple of the other (as was always the case in the present study), the difference between the two sample means, $\bar{X}_1 - \bar{X}_2$, may be regarded as the mean of N values of a single variable, $V = \frac{1}{C_1} \sum_1^{C_1} X_1 - \frac{1}{C_2} \sum_1^{C_2} X_2 = \bar{X}_{1(C_1)} - \bar{X}_{2(C_2)}$, where C_1 and C_2 are integers one of which is 1. Therefore the Central Limit Theorem applies and consequently the general robustness of Z (i.e., the overall fit of Z to a normal distribution with the same mean variance and area) should increase monotonically with increasing N, provided that C_1 and C_2 remain constant, i.e., should increase monotonically with increasing absolute sample size provided relative sample size is held constant. Thus, two-sample Z assumes about the sampled populations only that they are normal, but when the assumption is violated, its robustness is influenced by at least the following factors:

(a) "degree" of nonnormality (especially the asymmetry aspect) of each of the sampled populations

(b) relative population shapes (especially, degree of similarity of shape when shapes are asymmetric, variances are homogeneous and sample sizes are equal)

(c) relative population variances

(d) absolute sample sizes

(e) relative sample sizes

(f) size of significance level

(g) location of rejection region.

The two-sample F test (identified as R in this study) for equal population vairances, $F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{\frac{\sum (X_1 - \bar{X}_1)^2}{N_1 - 1}}{\frac{\sum (X_2 - \bar{X}_2)^2}{N_2 - 1}}$ assumes (aside from independent random sampling) only that both sampled populations are normally distributed. When the sampled population is quite nonnormal, the distribution of $\hat{\sigma}^2$ tends to depart drastically from its normal-theory distribution. And the situation does not tend to improve much with increasing sample size. Thus the N's can have little ameliorative influence upon the shape of the distributions of the $\hat{\sigma}^2$'s although their relative sizes can weight the relative influence of the two shapes (since the variance of the distribution of $\hat{\sigma}^2$ decreases as N increases). Therefore if one population is quite nonnormal the only case in which F might be expected to prove fairly robust is that in which the other population is normal or quasi-normal and contributes a very much smaller sample. Since a ratio of two variance estimates is involved, relative population shape must be presumed to have at least some influence. Thus the two-sample F test for equal population variances assumes about the sampled populations only that they are both normally distributed, but when that assumption is violated its robustness is influenced by the following factors:

(a) "degree" of nonnormality of each of the sampled populations

(b) relative population shapes

(c) absolute sample sizes

(d) relative sample sizes

(e) size of significance level

(f) location of rejection region.

The one-sample t test, $t = \frac{\bar{X} - \mu}{\sqrt{\frac{\sum (X - \bar{X})^2}{N(N - 1)}}}$, assumes about the sampled population only that it is normally distributed. The formula can be written

$$t = \frac{Z\ \sigma/\sqrt{N}}{\sqrt{\frac{\sum (X - \bar{X})^2}{N\ (N - 1)}}} = \frac{Z\ \sigma/\sqrt{N}}{\sqrt{\frac{\hat{\sigma}^2}{N}}} = \frac{Z\ \sigma}{\sqrt{\hat{\sigma}^2}}$$

, i.e., a constant times the ratio of the one-sample Z statistic to the square root of the sample estimate of the population variance. The latter estimate is extremely nonrobust (at least in regard to the shape and variance of its distribution), is uncorrelated with $\bar{X}$ (and therefore with the numerator of t) only if the sampled population is normal, but is an unbiassed estimate of σ^2 regardless of whether the population is normal or otherwise. As sample size increases, the nonrobustness of t's numerator (in effect, Z) diminishes but that of its denominator does not (to any very appreciable degree). However, as N increases, the variance of the nonrobust denominator diminishes much faster than does the variance of t's increasingly robust numerator. Therefore, as sample size increases an increasingly robust numerator receives greater and greater weight, relative to a nonrobust denominator, in influencing the robustness of t. It is clear from the above formulas that all of the factors influencing the robustness of one-sample Z or of $\hat{\sigma}^2$ also influence the robustness of one-sample t. Also influential are factors associated with the relationship between numerator and denominator. Thus while the one-sample t statistic assumes about the sampled population only that it is normally distributed, its robustness, when that assumption is violated, is influenced by the following factors:

(a) "degree" of nonnormality (especially the asymmetric aspect) of the sampled population

(b) amount of resulting correlation between sample mean and sample variance

(c) size of sample

(d) size of significance level

(e) location of rejection region.

The two-sample t statistic, $t = \dfrac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sum (X_1 - \bar{X}_1)^2 + \sum (X_2 - \bar{X}_2)^2}{N_1 + N_2 - 2} \left(\dfrac{1}{N_1} + \dfrac{1}{N_2}\right)}}$

assumes (in addition to independent random sampling) that both sampled populations are normally distributed and that they have identical variances. The formula can be written as

$$t = \frac{\dfrac{Z\sqrt{\dfrac{\sigma_1^{\,2}}{N_1} + \dfrac{\sigma_2^{\,2}}{N_2}}}{}}{\sqrt{\dfrac{(N_1 - 1)\,\hat{\sigma}_1^{\,2} + (N_2 - 1)\,\hat{\sigma}_2^{\,2}}{N_1 + N_2 - 2}\left(\dfrac{1}{N_1} + \dfrac{1}{N_2}\right)}}$$

$$= \frac{Z\,\sigma_{\bar{X}_1 - \bar{X}_2}}{\sqrt{\dfrac{(N_1 - 1)\,\hat{\sigma}_1^{\,2} + (N_2 - 1)\,\hat{\sigma}_2^{\,2}}{N_1 + N_2 - 2}\left(\dfrac{1}{N_1} + \dfrac{1}{N_2}\right)}}$$

where Z is the two-sample Z statistic and, of course, $\sigma_{\bar{X}_1 - \bar{X}_2}$ is a constant. Therefore, the robustness of the numerator is influenced by all of the previously-listed factors which affect the robustness of the two-sample Z statistic. However, it is the denominator which is most vulnerable. Heterogeneity of variance causes the locations, i.e., means, of the distributions of $\hat{\sigma}_1^{\,2}$ and $\hat{\sigma}_2^{\,2}$ to be different, whereas under normal-theory they must be identical. Nonnormality does not bias the location but it greatly distorts the shape and variance of a $\hat{\sigma}^2$'s distribution. Furthermore, although they are independent under normal theory, nonnormality causes the $\hat{\sigma}^2$ and $\bar{X}$ calculated from the same sample to be correlated. As the size, N_i, of a given sample increases, the mean of the distribution of $\hat{\sigma}_i^{\,2}$ remains constant but the variance diminishes and, therefore, nonnormality which distorts the latter tends to become decreasingly important (in influencing the robustness of the denominator) relative to heterogeneity which exerts its influence through the former.

In contributing to the denominator, each $\hat{\sigma}_i^{\,2}$ is weighted by $N_i - 1$, thus permitting the $\hat{\sigma}_i^{\,2}$ to influence the denominator roughly in proportion to relative sample size. Furthermore, even if both $\hat{\sigma}_1^{\,2}$ and $\hat{\sigma}_2^{\,2}$ are extremely nonrobust, the weighted averaging process affords the opportunity for extreme values to be somewhat moderated by combination with fortuitously less extreme values, or offset by combination with oppositely extreme values in their net effect upon the robustness of the entire denominator. Likewise, the correlations between $\hat{\sigma}_i^2$ and $\bar{X}_i$ for samples from dissimilar populations may tend to average out in producing the net correlation between numerator and denominator. (However, the correlations may have been beneficial, rather than detrimental, to robustness.)

Only if variances are homogeneous or if samples are of equal size does the squared denominator (i.e., the expression under the radical) reduce to $\hat{\sigma}^2_{\bar{X}_1-\bar{X}_2}$, i.e., an unbiassed estimate of the variance of the numerator, which it was intended to be, and is under normal theory. When only the normality assumption is violated (or when variances are heterogeneous but samples are of equal size), the distribution of the squared denominator has its normal theory mean, but not its normal-theory shape or variance. When variances are heterogeneous and sample sizes are unequal, it has none of these normal-theory characteristics.

As absolute sample size increases (i.e., as N_1 and N_2 increase simultaneously and proportionately) the numerator, which contains only a difference between means, becomes increasingly robust due to the Central Limit effect and, for the same reason, becomes completely robust when absolute sample size becomes infinite. (We are regarding the homogeneity assumption as being required by the denominator alone, since only the denominator differentiates the t test from the Z test, which does not require that variances be homogeneous.) The shape and variance of the denominator's distribution are quite nonrobust against either nonnormality or heterogeneity at all sample sizes. However, this eventually becomes academic as absolute sample size increases because the variance of the denominator diminishes both absolutely and relative to that of the numerator, until at infinite sample size the variance of the denominator about the mean of its distribution has shrunk to nil. At this point the denominator is a functional constant and can be represented by the mean of its distribution. But that mean has its normal-theory value only when variances are homogeneous or samples are of equal size. Therefore, even when sample sizes are infinite, the two-sample t test is completely robust only against nonnormality in general; in the special case where sample sizes are equal, it is also robust against heterogeneity of variance. If $\sigma_2^2/\sigma_1^2 = R$, $N_2/N_1 = r$, and sample sizes are infinite, the ratio of the actual standard deviation of the t distribution to the standard deviation it would have under normal theory (i.e., when $R = 1$) is

$$\frac{\text{Ultimate Standard Deviation of t}}{\text{Ultimate Normal-Theory Standard Deviation of t}} = \sqrt{\frac{r + R}{rR + 1}} .$$ This ratio departs monotonically from 1 as either r or R departs monotonically from 1 (while the other variable is held constant at a value not equal to unity). For a constant value of r exceeding 1, the ratio reaches a maximum of $\sqrt{r}$ when $R = 0$ and a minimum of $1/\sqrt{r}$ when $R = \infty$. Therefore, for unspecified values of r and R the ratio has an upper bound of infinity and a lower bound of zero. Thus, the unqualified statement that the two-sample t test is robust may be "infinitely" in error. And, in a more practical vein, the ratio shows that the test cannot achieve complete robustness in any case where r and R are both different from 1. So when both population variances and sample sizes are unequal, the variance of the t distribution is distorted, even when sample sizes are infinite, and, as a result, $\bar{\rho}$ approaches a limiting value different from α.

Thus the two-sample t test assumes about the sampled populations only that they are normally distributed with equal variances, but when assumptions are violated, its robustness is influenced by the following factors:

(a) "degree" of nonnormality (especially the asymmetry aspect) of each of the sampled populations

(b) relative population variances

(c) relative population shapes

(d) absolute amount of correlation between sample mean and sample variance, for each population

(e) relative amounts (and directions) of correlation between sample mean and sample variance

(f) absolute sample sizes

(g) relative sample sizes

(h) size of significance level

(i) location of rejection region.

The analysis-of-variance F test,

$$F = \frac{\dfrac{\sum_1^k N_i (\bar{X}_i - \bar{\bar{X}})^2}{k - 1}}{\dfrac{\sum_{i=1}^k \sum_{j=1}^{N_i} (X_{ij} - \bar{X}_i)^2}{\sum_1^k (N_i - 1)}} = \frac{\dfrac{\sum_1^k N_i (\bar{X}_i - \bar{\bar{X}})^2}{k - 1}}{\dfrac{\sum_{i=1}^k (N_i - 1)\hat{\sigma}_i^2}{\sum_1^k (N_i - 1)}}$$

assumes (aside from independent random sampling) that all sampled populations are normally distributed and have the same variance. In the two sample case, this F test is equivalent to a two-tailed, two-sample t test. When there are more than two samples, it is, in a sense, a multi-sample analogue of the two-tailed, two-sample t test, and its robustness is affected by much the same factors as affect the robustness of that test. When there are more than two samples, its denominator is the average (i.e., the weighted sum) of a greater number of variance estimates, thus permitting more effective compensation for extreme values by amalgamation with moderate (or opposite-extreme) values, and also facilitating the attenuation or dissolution of correlation between numerator and denominator, due to correlation between sample means and sample variances when samples come from nonidentical populations. (The individual correlations may have been beneficial, rather than detrimental, to robustness, however, so that their mutual cancellations, etc., cause $\bar{\rho}$ to depart further from α than when they operated in undiminished force.)

Only if variances are homogeneous or if sample sizes are equal do the distributions of the numerator and of the denominator have the same mean, as they must under normal theory. (And only in the former case is that common mean the same value as the common mean under normal theory.) Thus inequality of both population variances and sample sizes causes a shift in the location of the numerator's distribution relative to that of the denominator's, and this tends seriously to bias the location of the F distribution with consequent deviations of $\bar{\rho}$ from α. The shapes and variances of the distributions of numerator and denominator are distorted by violation of either the normality or homogeneity assumption. However, except for those produced in the numerator's distribution by heterogeneity of variance, such distortions of shape and variance diminish in importance with increasing N. As absolute sample size increases the sample means, of which the numerator is entirely composed, become more and more nearly normally distributed, due to the Central Limit effect, and the distortions associated with their diminishing nonnormality diminish also and vanish when infinite sample sizes produce complete normality in the distributions of the means. Also, as absolute sample size increases, the variance of the denominator's distribution about its mean shrinks both absolutely and relative to the variance of the numerator until, at N = infinity, the denominator's distribution consists solely of its mean about which there is no longer any dispersion. The variance and shape of the denominator's distribution remain highly nonrobust at all sample sizes. However, distortions in a vanishing variance or in a virtually one-dimensional shape will contribute negligibly to the influence of the denominator upon the robustness of F. (See appendix VII)

Even at infinite sample size, however, heterogeneity of variance prevents the F distribution from having its normal-theory variance (except in the two-sample case where $F = t^2$, and in an "impractical" case involving an infinite number of samples) and, when samples are unequal in size, also prevents it from having its normal-theory mean. In either case, the result is that $\bar{\rho}$ approaches a limiting value different from α. Nonnormality however produces no distortion in the F distribution at infinite N. At infinite sample size the F test is completely robust against nonnormality alone, and its nonrobustness against heterogeneity remains unaltered by the additional violation of the normality assumption. Some idea of the robustness of F, at infinite sample size, when only the variance of its distribution is distorted can be gotten from the following. If F is based upon K samples, each of size N from a population with variance σ^2 and rK samples, each also of size N, from a population with the same mean but variance $R\sigma^2$, then when N becomes infinite the F distribution has its normal theory mean and a standard deviation whose ratio to the normal-theory standard deviation of F is $\sqrt{1 + r\left(\frac{rK + K - 2}{rK + K - 1}\right)\left(\frac{R - 1}{rR + 1}\right)^2}$. This ratio has a minimum value of 1 which it reaches under any of three conditions: (a) $R = 1$, (b) $K = 1$ and $r = 1$, in which case there are only two samples and $F = t^2$, (c) r = infinity. If r and R are held constant, the ratio increases monotonically with increasing K from $\sqrt{1 + (r - 1)\left(\frac{R - 1}{rR + 1}\right)^2}$ at K = 1 to $\sqrt{1 + r\left(\frac{R - 1}{rR + 1}\right)^2}$ at K = infinity. If r and K are held constant the ratio increases monotonically (a) as R increases from 1 (at which the ratio is 1) to infinity, at which the ratio is $\sqrt{1 + \frac{r - 1}{r^2}}$ for K = 1 and is $\sqrt{1 + \frac{1}{r}}$ for K = infinity, (b) as R decreases from 1 to zero, at which the ratio is $\sqrt{r}$ for K = 1 and $\sqrt{r + 1}$ for K = infinity. Finally, if R and K are held constant, the ratio increases monotonically as r increases from zero (actually 1/K is the smallest meaningful value that r can assume) to a critical value which depends upon R and K, after which it decreases monotonically with further increases in r, eventually descending to the "completely robust" value of 1 when r becomes infinite. The critical value is (2R + 1)/R for K = 1 and 1/R for K = infinity. For $r > 1$, the ratio has its maximum value at R = 0, where it becomes $\sqrt{r}$ for K = 1 and $\sqrt{r + 1}$ for K = infinity. Since the critical value of r at R = 0 is infinity for either K = 1 or K = infinity, the maximum value of the ratio, for unspecified values of K, r and R, is infinity. Thus, even if sample sizes are made equal and infinite in order to give the F distribution its normal-theory mean and immunize it to nonnormality, the possible nonrobustness of F is still unlimited (without the introduction of further qualification).

So, although the F test assumes about the sampled populations only that they have normal distributions and homogeneous variances, when assumptions are violated its robustness is influenced by the following factors:

(a) "degree" of nonnormality of each of the sampled populations

(b) relative population shapes

(c) relative population variances

(d) absolute amount of correlation between sample mean and variance for each population

(e) relative amounts (and directions) of correlation between sample mean and sample variance

(f) absolute sample sizes

(g) relative sample sizes

(h) size of significance level.

Not only are there many factors contributing to nonrobustness, but a single factor may produce more than one effect. For example, nonnormality affects the distributions of both the sample mean and the sample estimate of population variance. Thus it affects both numerator and denominator of t or of F, but not necessarily in the same way. A long positive skew is conducive to both extremely large sample means and extremely large sample variances. The former tend to make one-sample t "too large" and the latter tend to counteract the former and make t "too central", the latter predominating, so that $\bar{\rho} < \alpha$ at the right tail. But samples drawn from the other end of the distribution tend to have means which are smaller than generally expected and variances which are extremely small, relative to normal theory expectations. The former tend to make one-sample t "moderately small" and the latter tend to augment the former and make t "much too small", the latter predominating so that $\bar{\rho} > \alpha$ at the left tail. Thus the single factor of nonnormality produces two different effects which may operate in the same or opposite directions, depending upon which end of the distribution of t one is interested in, i.e., depending upon a second factor, location of rejection region. Again, if one holds constant the size of one sample while increasing that of the other sample (all other factors held constant), the weight with which the increased sample's mean influences the robustness of the numerator of two-sample t will decrease but the weight with which that sample's estimate of population variance influences t's denominator will increase. If the increased sample was drawn from a normal population and the constant sample from a nonnormal population, the denominator of t will tend to become more robust but its numerator will tend to become less robust. The opposite will be the case if the populations are switched. And if both populations are nonnormal, the robustness of either numerator or denominator may be either improved or reduced, depending upon the original relative nonrobustness and relative weight of the increased sample's mean and estimate of population variance. These, in turn, depend upon relative population shape, relative population variance and absolute sample sizes, but not necessarily to equal degrees for numerator and denominator. Thus the single factor of relative sample size affects the influence, i.e., weight, of sample mean and variance upon the robustness of t in opposite directions; if one is increased the other is diminished. But whether the increase or decrease of influence is good or bad depends upon initial conditions involving additional factors such as relative population shapes and variances and absolute sample sizes. It is clear therefore that not only does nonrobustness have multiple causes, but also that its causes have multiple effects.

It has been shown, therefore, that robustness is influenced by a host of factors other than the simple violation of assumptions and that these factors do not necessarily act in concert nor with constant force or direction in producing the net effect obtained. There are "cancellations", "summations" and "augmentations" or interactions among the influential factors with sometimes curious results such as a drastic decrease in robustness with increasing absolute sample size. When variances are heterogeneous, the true significance level may not approach the nominal significance level as a limiting value, and the test may be even less robust when sample sizes are infinite than it is when they are very small. In short, there is practically no general statement that can be made about robustness to which there are no exceptions. The safest such statement is that robustness is a dangerous concept.

CONCLUSIONS

Statements such as "The t-test is robust against violation of the normality assumption." are quite common. (Other statistics may be substituted for t and other assumptions for normality, in the above.) However, such statements are extremely misleading and dangerous for the following reasons:

(a) There is no commonly accepted definition of what constitutes robustness, no agreed-upon criterion which distinguishes between a condition of robustness and one of nonrobustness. Therefore while the quoted statement may mean something to the utterer, provided he has an implicit subjective criterion of robustness, it can be (validly) meaningful to the hearer only if he is a mind reader.

(b) The statement does not quantify or otherwise specify the extent to which the assumption is violated. However (other things being equal) the greater the extent or degree of a single violation the farther the true significance level tends to depart from the nominal one, and often the rate is such that the departure will eventually exceed any reasonable fixed criterion of "robustness", thereby negating the statement. If we assume that the statement implicitly refers to "ordinary" violations of assumptions, we are back at the semantic impasse encountered in (a) above, since there is no commonly accepted criterion separating "ordinary" from "extraordinary" degrees of violation.

(c) When an assumption is violated, the discrepancy between true and nominal significance levels is not a simple function of the "degree" to which the assumption was violated. Instead, it is influenced by a multitude of additional factors which are not involved in the statement of the assumption (and which do not appear in the quoted statement claiming robustness) but which interact with the violation when it occurs. Such factors are:

(1) size of nominal significance level

(2) location of rejection region

(3) absolute sample sizes

(4) relative sample sizes

(5) relative sizes of population variances (when homogeneity of variance is not an assumption)

(6) relative shapes of the sampled populations

(7) absolute correlation between sample means and variances for each population

(8) relative correlations between sample means and variances among the various populations sampled.

The interactions are likely to be exceedingly complex and of high order. That is, the extent to which a given factor influences robustness generally depends not only upon its own value, but also upon the particular combination of factors involved and the value of each, and the interdependency is often quite strong (see below).

(d) When more than a single factor is called into play by assumption-violation, the factors may operate in different directions (some may tend to cause the true significance level to exceed the nominal one, while others may tend to cause the reverse) and may vary in relative potency as absolute sample size increases (one factor perhaps being prepotent at one sample size, giving way to another factor at a larger sample size). As a net result, the absolute discrepancy between true and nominal significance levels may not diminish monotonically as absolute sample size increases and, in fact, the discrepancy may assume both algebraic signs, the true significance level greatly exceeding the nominal level at small sample sizes, nearly equalling it at some intermediate sample size, and being greatly exceeded by it at large absolute sample sizes, or vice versa. Therefore "robustness" may actually decrease as a consequence of taking more data (i.e., as the absolute sample sizes increase while relative sizes remain fixed; it may also happen as a result of increasing the size of one sample while holding constant the size of the other samples.)

(e) Violation of certain assumptions may cause combined-sample estimates of population parameters to be biased (or may bias certain components of the test statistic, such as its denominator) with the result that the limiting value (when absolute sample sizes are all infinitely large) of the true significance level is quite different from the value of the nominal significance level. In other words even if based upon an infinite amount of data the test statistic may have a true distribution which differs greatly from its distribution under normal theory.

(f) From the above considerations it is clear that in order to convert the quoted statement into a meaningful and accurate statement about robustness it would have to be accompanied by a quantitative definition of robustness, a complete quantitative statement of the degree or extent of the violation, and a complete specification of the exact sampling and test conditions under which the test is to be performed. If all this were done, the statement would be so particularistic as to have little general appeal, which perhaps explains why the type of statement quoted survives in its amorphous undefined, unqualified form. It also explains why that form is so completely inaccurate and so utterly meaningless.

BIBLIOGRAPHY*

1. Bradley, James V., Studies in Research Methodology: I. Compatibility of Psychological Measurements with Parametric Assumptions, Wright Air Development Center Technical Report 58-574(I), September 1959.

2. Bradley, James V., Studies in Research Methodology: II. Consequences of Violating Parametric Assumptions - Fact and Fallacy, Wright Air Development Center Technical Report 58-574(II), September 1959.

3. Bradley, James V., Studies in Research Methodology: III. Persistence of Sequential Effects Despite Extended Practice, MRL Technical Documentary Report 62-60, 6570th Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, June 1962.

4. Bradley, James V., Studies in Research Methodology: IV. A Sampling Study of the Central Limit Theorem and the Robustness of One-Sample Parametric Tests, AMRL Technical Documentary Report 63-29, 6570th Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, March 1963.

5. Bradley, James V., Studies in Research Methodology: V. Irrelevance in the t-Test's Rejection Region, AMRL Technical Documentary Report 63-109, 6570th Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, December 1963.

6. Burington, Richard S. & May, Donald C., Handbook of Probability and Statistics with Tables, Sandusky, Ohio:Handbook Publishers, Inc., 1953, pp 22-23, 83, 135, 141-142, 147, 151, 154-156, 158-159, 203-204, 243.

7. Federighi, Enrico T., "Extended Tables of the Percentage Points of Student's t-Distribution," J. Amer. Statist. Assn., 1959, 54, 683-688.

8. Feller, William, An Introduction to Probability Theory and Its Applications Vol I, 2nd Ed., New York:Wiley, 1957, page 64.

9. Fisher, Ronald A. & Cornish, E. A., "The Percentile Points of Distributions Having Known Cumulants," Technometrics, 1960, 2, 209-225.

10. Fisher, Ronald A. & Yates, Frank, Statistical Tables for Biological, Agricultural and Medical Research, 3rd Ed., New York:Hafner, 1949, pp. 31, 35, 43.

*For a bibliography of studies of the robustness of Z, t and F see reference 4 above.

11. Glover, James W., Tables of Applied Mathematics in Finance, Insurance, Statistics, Ann Arbor, Mich.:George Wahr, Publisher, 1923, pp. 482-486.

12. Guilford, J. P., Psychometric Methods, New York:McGraw-Hill, 1954, pp. 559-562.

13. Hald, A., Statistical Tables and Formulas, New York:Wiley, 1952, pp. 47-59.

14. Kelley, Truman L., The Kelley Statistical Tables, Cambridge, Mass.:Harvard Univ. Press, 1948, pp. 36-137.

15. Mood, Alexander M., Introduction to the Theory of Statistics, New York: McGraw-Hill, 1950, p. 16.

16. Owen, Donald B., Handbook of Statistical Tables, Reading, Mass.:Addison-Wesley, 1962, pp. 28-30, 63-87.

17. Salzer, Herbert E., Tables of $n!$ and $\Gamma(n + 1/2)$ for the First Thousand Values of n, National Bureau of Standards Applied Mathematics Series 16, Washington, D. C.:U. S. Govt. Printing Office, 1951.

18. Spenceley, George W., Spenceley, Rheba M. & Epperson, Eugene R., Smithsonian Logarithmic Tables, Washington, D. C.:Smithsonian Institution, 1952.

19. Wilks, Samuel S., Mathematical Statistics, New York:Wiley, 1962, pp. 82-83, 183-187, 199.

APPENDIX I

Formulas Used to Calculate Test Statistics

The body of this report gave the formulas for the test statistics in the notation usually found in textbooks and did so without definition of the symbols used. In this appendix the formulas actually used will be given in the notation used in this report and all symbols and operations will be explicitly defined. Each test statistic is identified by the code which was explained earlier (and it is the "Appendix" form of the code which is used here). First the terms required in the formulas are defined as follows. These terms, and the operations which produce them have a logic which is based upon Table I and (the right hand sides of) Tables II and III.

i = An integer subscript where $1 \leq i \leq 4N$.

P_i = An observation drawn from population P by the i th random number, in order of sequence, in the fundamental block or pool of 4N random numbers.

Q_i = An observation drawn from population Q by the i th random number.

$D_i = X_i - Y_i$ where the subscripts of X and Y are identical.

$$\bar{P}_1 = \frac{\sum_{i=1}^{N} P_i}{N} \qquad SS(P_1) = \sum_{i=1}^{N} (P_i - \bar{P}_1)^2$$

$$\bar{P}_2 = \frac{\sum_{i=N+1}^{2N} P_i}{N} \qquad SS(P_2) = \sum_{i=N+1}^{2N} (P_i - \bar{P}_2)^2$$

$$\bar{P}_3 = \frac{\sum_{i=2N+1}^{3N} P_i}{N} \qquad SS(P_3) = \sum_{i=2N+1}^{3N} (P_i - \bar{P}_3)^2$$

$$\bar{P}_4 = \frac{\sum_{i=3N+1}^{4N} P_i}{N} \qquad SS(P_4) = \sum_{i=3N+1}^{4N} (P_i - \bar{P}_4)^2$$

$$\bar{Q}_4 = \frac{\sum_{i=3N+1}^{4N} Q_i}{N} \qquad SS(Q_4) = \sum_{i=3N+1}^{4N} (Q_i - \bar{Q}_4)^2$$

$$\bar{P}_{12} = \frac{\sum_{i=1}^{2N} P_i}{2N} \qquad SS(P_{12}) = \sum_{i=1}^{2N} (P_i - \bar{P}_{12})^2$$

$$\bar{P}_{123} = \frac{\sum_{i=1}^{3N} P_i}{3N} \qquad SS(P_{123}) = \sum_{i=1}^{3N} (P_i - \bar{P}_{123})^2$$

$$\bar{P}_{1234} = \frac{\sum_{i=1}^{4N} P_i}{4N} \qquad SS(P_{1234}) = \sum_{i=1}^{4N} (P_i - \bar{P}_{1234})^2$$

$$\bar{Q}_{34} = \frac{\sum_{i=2N+1}^{4N} Q_i}{2N} \qquad SS(Q_{34}) = \sum_{i=2N+1}^{4N} (Q_i - \bar{Q}_{34})^2$$

$$\bar{D}_{14} = \frac{\sum_{i=1}^{N} D_i + \sum_{i=3N+1}^{4N} D_i}{2N} \qquad SS(D_{14}) = \sum_{i=1}^{N} (D_i - \bar{D}_{14})^2 + \sum_{i=3N+1}^{4N} (D_i - \bar{D}_{14})^2$$

G = Grand mean of all observations actually used in a test.

Next, the formulas actually used to calculate the test statistics are as follows:

$$ZP\ K = \frac{\bar{P}_K - 100}{\sigma_P/\sqrt{N}}$$

$$ZP\ 1234 = \frac{\bar{P}_{1234} - 100}{\sigma_P/\sqrt{4N}}$$

$$TP\ K = \frac{\bar{P}_K - 100}{\sqrt{\frac{SS(P_K)}{N(N-1)}}} \qquad df = N - 1$$

$$TP\ 1234 = \frac{\bar{P}_{1234} - 100}{\sqrt{\frac{SS(P_{1234})}{4N(4N-1)}}} \qquad df = 4N - 1$$

$$TD\ 2N = \frac{\bar{D}_{14}}{\sqrt{\frac{SS(D_{14})}{2N(2N-1)}}} \qquad df = 2N - 1$$

$$ZPQ\ C_1NC_2N = \frac{\bar{P}_{1....C_1} - \bar{Q}_{(4-C_2+1)---4}}{\sqrt{\frac{\sigma_P^2}{C_1N} + \frac{\sigma_Q^2}{C_2N}}}$$

$$TPQ\ C_1NC_2N = \frac{\bar{P}_{1....C_1} - \bar{Q}_{(4-C_2+1)---4}}{\sqrt{\frac{SS(P_{1....C_1}) + SS(Q_{(4-C_2+1)---4})}{C_1N + C_2N - 2}\left(\frac{1}{C_1N} + \frac{1}{C_2N}\right)}} \qquad df = C_1N + C_2N - 2$$

$$RPQ\ 2N2N = \frac{SS(P_{12})}{SS(Q_{34})} \qquad df = 2N - 1,\ 2N - 1$$

$$FP\ 2NNN = \frac{\frac{2(\bar{P}_{12} - \bar{P}_{1234})^2 + (\bar{P}_3 - \bar{P}_{1234})^2 + (\bar{P}_4 - \bar{P}_{1234})^2}{2}}{\frac{SS(P_{12}) + SS(P_3) + SS(P_4)}{N(4N-3)}} \qquad df = 2,\ 4N - 3$$

$$\text{FPQ } NC_2NC_3NN = \frac{\dfrac{(\overline{P}_1 - G)^2 + C_2(\overline{P}_2 - G)^2 + C_3(\overline{P}_3 - G)^2 + (\overline{P}_4 - G)^2}{1 + C_2 + C_3}}{\dfrac{SS(P_1) + C_2[SS(P_2)] + C_3[SS(P_3)] + SS(P_4)}{N(N-1)\ (2+C_2+C_3)}} \qquad df = 1 + C_2 + C_3,\ (N-1)\ (2+C_2+C_3)$$

$$\text{where } G = \frac{\overline{P}_1 + C_2\,\overline{P}_2 + C_3\,\overline{P}_3 + \overline{P}_4}{C_1 + C_2 + C_3 + C_4}$$

$C_2 = 0$ for FPQ NONN

$C_3 = 0$ for FPQ NNON

$C_i = 1$ otherwise

Some slight algebraic manipulations and occasional substitution of equivalent operations were sometimes performed on these formulas in programming them for the computer. However, the net effect was always the same as that which obtains above.

APPENDIX II

Sampling

Sampling was performed by an IBM 7090 computer, as were the calculations and tabulations required. Random numbers (different from those used in the earlier study and ultimately reduced to five digits each) were obtained by means of a pseudo-random number generating subroutine. The method was essentially the same as that used in the latter part of the previous study, but different in certain particulars. The generating formula was

$$R_i = \lambda R_{i-1} \qquad (\text{modulo } 2^{35}) \qquad i = 1, 2, ---N$$

where R is the random number and λ is a constant multiplier. The starting number R_o was changed after every N random numbers according to the formula

$$R_o = B_j \qquad j = 1, 2, 3, ----------2^{35} \quad \text{where}$$

A_o	$= a$	$= F_o\ a$
B_o	$= b$	$= F_o\ b$
$A_1 = A_o + B_o$	$= a + b$	$= F_o\ a + F_1\ b$
$B_1 = A_1 + B_o$	$= a + 2b$	$= F_1\ a + F_2\ b$
$A_2 = A_1 + B_1$	$= 2a + 3b$	$= F_2\ a + F_3\ b$
$B_2 = A_2 + B_1$	$= 3a + 5b$	$= F_3\ a + F_4\ b$
$A_3 = A_2 + B_2$	$= 5a + 8b$	$= F_4\ a + F_5\ b$
$B_3 = A_3 + B_2$	$= 8a + 13b$	$= F_5\ a + F_6\ b$
. .		
. .		
. .		
$A_j = A_{j-1} + B_{j-1}$	=	$F_{2j-2}\ a + F_{2j-1}\ b$
$B_j = A_j + B_{j-1}$	=	$F_{2j-1}\ a + F_{2j}\ b$

where the F_k are the numbers of the Fibonacci series, 1, 1, 2, 3, 5, 8, 13, 21, 34, ... F_k ... where $F_k = F_{k-1} + F_{k-2}$.

Using the above formula with the following initial values, $\lambda = 5^{15}$, A_0 = 6,975,917,870, B_0 = 6,274,843,075, thirteen batches of 10,000(4N) random numbers were generated, the last values of A and B in one batch being used as the initial values of A and B for the succeeding batch. To each batch, there corresponded a single value of N, which therefore serves to identify the batch. The batches were run in the following sequence: N = 8, N = 4, N = 8, N = 16, N = 32, N = 64, N = 128, N = 2, N = 256, N = 512, N = 1024, N = 8, N = 2. The first batch for N = 8 and the first batch for N = 2 are the ones reported in this study. The others were run as a check on certain peculiar effects at N = 8 and N = 2. The "replications" demonstrated that the peculiar effects were real and not attributable to nonrandomness. Having served their purpose, the second and third batches for N = 8 and the second batch for N = 2 were discarded, leaving ten batches, one for each different value of N.

A predesignated five digits of each random number was used to select observations. A 1 was added to each five digit random number, creating a new number which ranged from 00001 to 100,000 rather than 00000 to 99,999, i.e., having the same range of values as the size ranks of the 100,000 units comprising the X or Y population if tied-for ranks were arbitrarily assigned to equal values. The new number therefore identified the X observation and the Y observation having this new number as size rank in their respective populations and these observations were then drawn. Thus if the five digit random number is j, the observations to be drawn are the X observation and Y observation having in their respective populations a cumulative probability of (j + 1)/100,000. This, of course, produces a measure of correlation between X observations and Y observations drawn with the same random numbers (See the earlier report for scatter diagrams reflecting this correlation) and this fact is used to produce correlated observations for the matched-pair t-test, TD 2N, where $D_i = X_i - Y_i$, and X_i and Y_i are drawn with the same random number.

Having sampled from the X and Y populations, there is no need actually to sample, i.e., draw individual observations, from the A and B populations. All of the required A and B population parameters and sample statistics can be obtained by linear transformation from the X and Y population parameters and sample statistics based on the X and Y observations already drawn. Thus $\mu_A = \mu_B = \mu_X = \mu_Y = \mu = 100$, $\sigma_A^2 = \sigma_B^2 = .25\sigma^2_X = .25\sigma^2_Y = .25\sigma^2$, $\bar{A} = 100 + .5(\bar{X} - 100)$, $\bar{B} = 100 + .5(\bar{Y} - 100)$, $\sum(A - \bar{A})^2 = .25\sum(X - \bar{X})^2$, $\sum(B - \bar{B})^2 = .25\sum(Y - \bar{Y})^2$.

For each 4N random numbers the machine drew 4N X-observations and 4N Y-observations, and calculated all of the X-sample,. Y-sample, A-sample and B-sample (and "D-sample") component statistics required (such as means or variances) and then used them to calculate a single value of each of the 122 test statistics investigated in this study. Therefore each batch of 10,000(4N) random numbers resulted in 122 empirical sampling distributions, each based on 10,000 values of a test statistic. Use of a common set of random numbers makes many of these sampling distributions "nonindependent"; however, this is regarded as a virtue since it has the effect of diminishing the relative effects of chance and therefore of increasing the comparability of one distribution with another at the same value of N.

For each test statistic's empirical frequency distribution, the machine cumulated frequencies over the left tail and printed out the empirical cumulative probabilities corresponding to values of the test statistic having normal-theory left-tail cumulative probabilities, α's, of .0005, .001, .005, .01, .025, .05, .10, .20, .30, .40 and .50. The analogous procedure was then followed for the right tail, the print-out format being the mirror image of that for the left tail. The cumulation included values falling exactly on the boundary value, so, as in the earlier study, the left and right tail cumulative frequencies at $\alpha = .50$ sum to more than 10,000 in the case of Z's and t's, the surplus being the number of cases of Z = 0 or t = 0, which are counted twice.

The left and right tail α's (or, more properly, the normal-theory statistic-values associated with them) may be considered as boundaries which divide the sampling distribution into 22 intervals or "cells" (23 if one counts the point corresponding to $\alpha = .50$ as an interval). In addition to the cumulations mentioned in the preceding paragraph, the machine kept count of the number of statistic-values falling in each interval. Values falling on an interval boundary were assigned to the more tailward of the two intervals having it as a common boundary; in the case of $\alpha = .50$, half of the values falling on the boundary separating left and right tails were assigned to each tail. A chi-square test of fit, with 21 degrees of freedom, was then performed using 10,000 $(\alpha_i - \alpha_{i-1})$ as the expected frequency for the cell "bounded" by α_{i-1} and α_i, i.e., basing expected frequencies upon normal-theory. Thus, a chi-square test of goodness of fit to normal-theory was performed upon each of the 122 sampling distributions at every value of N. The resulting value of chi-square was printed out, to the right of the right-tail cumulative probabilities, for each test statistic. The primary purpose of these chi-square tests was to imply satisfactory randomness of sampling when passed by empirical distributions of test statistics all of whose samples were drawn from the Y population. In all other cases the chi-square test was performed simply because it was easy to include this additional information and because in such cases the value of chi-square can serve as a sort of index of grossness of departure from normal-theory expectation. It is emphasized, however, that in these additional cases no hypothesis is being tested.

The normal-theory Z values corresponding to the α's used were obtained from Kelley's tables (ref. 14) and were checked against tables published by Guilford (ref. 12), and Fisher and Yates (ref. 10). They were carried in the machine to an accuracy of at least eight significant figures. The normal-theory t values corresponding to the α's used were obtained mainly from five decimal place figures personally supplied by Dr. Donald B. Owen of the Sandia Corporation, who used an existing IBM machine program to solve the Cornish-Fisher approximation (ref. 9) for t's corresponding to the α's and degrees of freedom required. Owen also used exact formulae to obtain all of the t values required at degrees of freedom 31, 46, 62, 63, 94, 126, and 127, and sent the resulting figures along with those obtained by the approximation. The five decimal place t's obtained with the exact formulae agreed exactly, in every case, with the five decimal place t's obtained by the approximate formulae used at the same number of degrees of freedom. Therefore, it was deemed entirely safe to use the t values obtained by the approximation for all degrees of freedom above 31 (degrees of freedom extended to as high as 4095). Below 31 degrees of freedom, the approximate t values were checked against existing published tables. If the five decimal place approximate value, when rounded to the number of decimal places in which the exact value was expressed in the table, agreed exactly with the tabled value, then the five-decimal-place approximate value was used. If it disagreed, the tabled value was used. The exact tables against which the check was made were: Federighi (ref. 7) for α = .0005 and .001, Owen (ref. 16) for α's of .005, .01, .025, .05 and .10, and Hald (ref. 13) for α's of .20, .30, and .40 (at α = .50, of course, t is exactly zero at all degrees of freedom). Federighi's tables give t values to three decimal places (which at α's of .0005 and .001 meant that t was expressed in six significant figures at df = 1, five at df = 2 and 3, and four at df = 4, 6, and 7). Owen's published tables give t values to four decimal places, and Hald's tables, in the sections used, give t to three decimal places. The result of these comparisons was that the approximate (Cornish-Fisher) values were used in all except the following cases: Federighi's t values were used at α = .0005 and α = .001 for degrees of freedom of 1, 2, 3, 4, 6, and 7 (i.e., all of the df required below df = 10); Owen's published t values were used at α = .005 for df of 1, 2, 3, 4, 6, and 7, at α = .01 for df of 1, 2, 3, 4, 6, and 7, at α = .025 for df of 1, 2, 3, and 4, at α = .05 for df of 1, 2, and 3, and at α = .10 for df of 1 and 2; Hald's t values were used at α's of .20, .30, and .40 for 1 degree of freedom. The t values selected were carried in the machine to as many significant figures as were present in the value.

The normal-theory F values corresponding to the α's used were obtained (for all cases, i.e., for all required degrees of freedom) by solving for F the formula

$$\frac{\Gamma\left(\frac{n+d}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{d}{2}\right)}\, n^{\frac{n}{2}}\, d^{\frac{d}{2}} \int_0^F \frac{X^{\frac{n-2}{2}}}{(nX+d)^{\frac{n+d}{2}}}\, dX = \alpha$$

n and d are the number of degrees of freedom for the numerator and denominator respectively. The formula was obtained from (ref. 16 page 63). Methods for its solution were conceived and supplied by Mr. Edwin Godfrey. For the three-sample F test, n = 2, and for this case an exact solution is easily obtained:

$$\frac{\Gamma\left(\frac{2+d}{2}\right)}{\Gamma\left(\frac{2}{2}\right)\Gamma\left(\frac{d}{2}\right)} \; 2^{\frac{2}{2}} \, d^{\frac{d}{2}} \int_{o}^{F} \frac{X^{\frac{2-2}{2}}}{(2X+d)^{\frac{2+d}{2}}} \, dX = \alpha$$

Substituting $\frac{d}{2}\,\Gamma\left(\frac{d}{2}\right)$ for $\Gamma\left(\frac{2+d}{2}\right)$ and 1 for $\Gamma\left(\frac{2}{2}\right)$

[See page 243 of (ref. 6)],

$$\frac{\frac{d}{2}\,\Gamma\left(\frac{d}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \; 2d^{\frac{d}{2}} \int_{o}^{F} (2X+d)^{-\frac{2+d}{2}} \, dX = \alpha$$

$$d^{\frac{d+2}{2}} \left[\frac{(2X+d)^{-\frac{d}{2}}}{2\left(-\frac{d}{2}\right)} \right]_{o}^{F} = \alpha$$

$$d^{\frac{d+2}{2}} \left[-d^{-1} \; (2F+d)^{-\frac{d}{2}} + d^{-\left(\frac{d+2}{2}\right)} \right] = \alpha$$

$$-d^{\frac{d}{2}} \, (2F+d)^{-\frac{d}{2}} = \alpha - 1$$

$$2F + d = \left[\frac{d^{\frac{d}{2}}}{1-\alpha} \right]^{\frac{2}{d}} = d \, (1-\alpha)^{-\frac{2}{d}}$$

$$F = \frac{d(1-\alpha)^{-\frac{2}{d}} - d}{2} = \frac{d}{2} \left[(1-\alpha)^{-\frac{2}{d}} - 1 \right]$$

This formula for F was solved by the machine, and for each value of α and each pair of degrees of freedom to be used F was printed out to six, seven or eight decimal places (the latter obtaining when $F < 1$, and always containing at least five significant figures). Where possible, these were then checked against existing tables. A number of discrepancies were discovered, all except one being discrepancies of one unit in the rightmost figure given in the published table and being explanable as due to round-off error by the compiler of the table. The single exception was investigated further and it was determined that the published value was in error. It was concluded that the obtained F values were highly accurate; however, it is not implied that they are necessarily accurate to the number of significant figures printed out by the machine. For the four-sample F test where $n = 3$ and for the two-sample F test where $n = d$, the expression on the left of the integral sign (of the formula given at the beginning of this paragraph) was evaluated by using logarithms. All of the gamma terms are either of the form $\Gamma(g)$ or $\Gamma(g + 1/2)$ where g is an integer. In the former case $\log \Gamma(g)$ to seven decimal places was obtained directly from (ref. 11) when $1 \leqq g \leqq 1000$. In the latter case $\Gamma(g + 1/2)$ to eight significant figures was obtained when $1 \leqq g \leqq 1000$ from (ref. 17) and its logarithm was calculated to ten decimal places using the tables and factorizing method of (ref. 18). When g exceeded 1000, $\Gamma(g)$ and $\Gamma(g + 1/2)$ were converted to factorials (using the relationships given on page 243 of (ref. 6)), and (see ref. 15 or 8) Stirlings approximation, $k! \cong \sqrt{2\pi}\, k^{k+\frac{1}{2}}\, e^{-k+\frac{1}{12k}}$, was substituted for the factorial. Thus, whatever the size of g, the expression to the left of the integration sign could be evaluated with considerable accuracy by adding and subtracting logarithms all of which were accurate to at least seven decimal places when $g \leqq 1000$, and probably were of the same order of accuracy when $g > 1000$.

The numerical value $U = \dfrac{\Gamma\left(\frac{n+d}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{d}{2}\right)}\, n^{\frac{n}{2}}\, d^{\frac{d}{2}}$, for each pair of n and d values required, was fed into the computer which then, using a method of successive approximations, solved the equation

$$U \int_0^F \frac{X^{\frac{n-2}{2}}}{(nX + d)^{\frac{n+d}{2}}}\, dX = \alpha$$

for the F's corresponding to the required values of α and to the required pairs of degrees of freedom. These F's were printed out to the same number of decimal places as before (when $n = 2$) and always to at least six significant figures. When checked against published tables with five significant figures, the worst discrepancy was an error of five or six units in the fifth significant figure, and errors of this order or less could account for all discrepancies with published tables giving F to three significant figures. As a further check, machine-computed, approximate F values at $n = d = 3$ were substituted into a formula representing an exact analytic solution for α in terms of F when $n = d = 3$. The formula was then solved for α. All 21 of the α values, computed from the 21 F's (df = 3, 3) substituted into the formula, were correct (i.e., were the α's desired in this study) to within $\pm$ one unit in the sixth decimal place. Thus a "backward check on α" confirms the validity of the F values to the degree of approximation desired. These checks all took place at low degrees of freedom. At the higher degrees of freedom not to be found exactly in published tables, approximate F values were obtained for the desired pair of degrees of freedom by interpolation from the published values. These approximate F's were then compared with the corresponding F value obtained from the machine. Discrepancies were of about the order to be expected from interpolation error. Therefore, it was concluded that the F values obtained by the machine were sufficiently accurate for the purposes of this study. In running the problem proper, the necessary machine-obtained F values were fed into the machine as inputs using exactly the values originally obtained by the machine and using all of the significant figures originally obtained. Thus at low degree of freedom F-values known to conflict with published tables were used rather than the correct value. The advantage of this procedure was to maintain a more or less uniform degree of inaccuracy in interval boundaries from one batch to another, thereby maintaining comparability between batches. The advantage of using all significant figures can be seen from the closeness of the F values corresponding to right-tail α's of .0010 and .0005 at 2047 and 2047 degrees of freedom; these F's are 1.147198 and 1.158068 respectively. Even if the last four figures contribute nothing to the absolute accuracy of a single F, they are desirable if they contribute to the accuracy of one F relative to a neighboring F, i.e., to the accuracy of the difference between adjacent F's, and this will be the case if inaccuracy in absolute values is attributable in whole or in part to a constant bias. Using all significant figures exploited a possible advantage without invoking any penalty whatever.

APPENDIX III

Checks

Before making final runs, several single "iterations" were run and checked. An iteration consisted of generating a single block of 4N random numbers, using them to select observations, and calculating a single value of each of the test statistics to be investigated. The machine printed out the 4N random numbers, the observations selected by them, each of the sample statistics required in calculating the test statistics, the machine-calculated values of the test statistic, and the tables presenting the cumulative frequency distributions of the test statistics (which in this case "cumulated" the single occurrence of the one calculated value of each test statistic). Thus it was possible to check every step of the arithmetic, searching, and tabulating operations performed by the machine, by using a desk calculator and referring to input tables. Several such checks were performed by the programmer and by the writer. A single error was discovered and corrected. As a further check, a print-out of the machine program was checked by the writer.

Along with the cumulative frequency distributions printed out in the final runs, the machine also printed out: (a) the input table giving, in effect, the X value and Y value having each of the size ranks (ties disallowed) from 1 to 100,000 in their respective populations, (b) the beginning and final values of A, B, and λ in generating the pseudo-random numbers used, (c) the normal-theory values of Z, t, and F corresponding to the α's used and having the degrees of freedom needed for the test statistics investigated at that value of N. The latter, (c), were thoroughly checked by the writer against the original normal-theory input tables submitted to the programmer (which included the F values generated and printed-out by the machine), and this was done for every batch, i.e., at every value of N. Also for every batch the writer checked the beginning values of A, B, and λ against their corresponding final values in the immediately preceding batch (λ, being a constant had the same value in every batch, and its "check" was redundant). For two widely separated batches the writer completely checked the XY input table mentioned in (a) above. Finally, the writer checked "by hand" several of the chi-square values printed-out by the machine for the normal theory tests of fit. All checks were passed.

The storage capacity of the machine, as used, was eight decimal-system digits for any single term. Consequently the maximum accuracy of which the machine was capable was eight significant figures. In the present problem its minimum accuracy can be expected to occur when calculating "sums of squares" by the formula

$$SS(X) = \sum_1^N X_i^2 - \frac{\left(\sum_1^N X_i\right)^2}{N}$$

when $\left(\sum_1^N X_i\right)^2/N$ is nearly equal in size to $\sum_1^N X_i^2$, i.e., when SS(X) assumes its smallest nonzero value. Because the X's and Y's are all integers, the smallest possible nonzero value of SS(X) or SS(Y) can be shown to be $1 - \frac{1}{N}$, and the smallest possible nonzero value of SS(A) or SS(B) is one-fourth of this. The smallest nonzero SS(X) therefore occurs at N = 2 and is .5. At N = 2 the largest value of $\sum_1^N X_i^2$ is a five digit integral number, so SS(X) will be carried in the machine to 8 - 5 = 3 decimal places. Therefore at N = 2, SS(X) will always be represented by at least three significant figures. Similar calculations based, not upon the smallest possible SS (which becomes increasingly improbable with increasing values of N), but rather upon the smallest value of SS actually obtained in the previous study (where all statistics were printed out) indicate that the smallest nonzero SS was always represented by at least three significant figures and, at some N's, by as many as six. Since these calculations are based upon the worst situation, it seems clear that the number of figures carried in the machine was always adequate for the needs of the problem. The conformity of the control distributions to the symmetrical or normal-theory distributions "predicted" for them supports the conclusion that sufficient accuracy was maintained at all N's and in all portions of the distributions.

As mentioned earlier, if sampling is sufficiently "random", then above N = 8 the statistics ZY 1, ZY 2, ZY 3, and ZY 4 should have distributions close enough to the normal-theory Z distribution to pass chi-square tests of fit to it. Each of these ZY statistics is based upon different and nonoverlapping sequences of the random numbers generated. Therefore they provide independent tests of randomness. Furthermore, each test checks the randomness of a sequence of random numbers which was used either in its entirety or not at all in the generation of the "sampling distribution" of one of the component samples upon which the one-, two-, three-, or four-sample test statistics are based. At $N \leq 8$, passing the chi-square test of fit to normal-theory still tends to support the inference of "sufficient randomness", but failing it does not necessarily impugn randomness since it is known from the earlier study that, at the smallest N values, the nonnormal discreteness of the Y population causes departures of ZY from its normal-theory distribution which are sometimes appreciable, i.e., detected by the chi-square test of fit to normality. Therefore, when the chi-square test of normality is failed at $N \leq 8$ it is appropriate to resort to a second test which is sensitive to nonrandomness but unaffected by discreteness. Since the Y population is exactly symmetrical, the true distributions of $\bar{Y}$ and ZY will be symmetrical also, irrespective of discreteness. Therefore, the supplementary test will be a test of symmetry on ZY. Let f_{Li} and f_{Ri} be the frequencies with which empirical ZY values fell in the i-th interval to the left of $\alpha = .50$, and the i-th interval to the right of $\alpha = .50$, respectively, and let the expected frequency f_{ei} be their average $\frac{f_{Li} + f_{Ri}}{2}$.

Then $\chi_i^2 = \frac{(f_{Li} - f_{ei})^2}{f_{ei}} + \frac{(f_{Ri} - f_{ei})^2}{f_{ei}} = \frac{2(f_{Li} - f_{ei})^2}{f_{ei}}$ is a chi-square test,

with one degree of freedom, of the hypothesis that the true expected frequencies for the two cells are of equal (but otherwise unspecified) value, and

$\chi^2 = \sum_{i=1}^{11} \chi_i^2$ is a chi-square test, with 11 degrees of freedom, of the hypothesis

that the true histogram of Z_Y, with each of the 22 intervals used in the test represented by a single histogram bar, is exactly symmetrical. (In this test values of ZY = 0 are discarded, rather than divided equally between the adjoining intervals.)

It is these tests on individual ZY distributions upon which the criterion of acceptable randomness is based. Results of these, and other, tests are shown in Table VII.

TABLE VII

Results of Chi-Square Tests on Control Distributions

(Cell entries are values of χ^2. One, two, and three underlines indicate significance at .05, .01, and .001 levels, respectively.)

Goodness-of-Fit-to-Normality Tests (21 degrees of freedom)

N	Control Distribution: ZY 1	ZY 2	ZY 3	ZY 4	ZY 1234	RY 2N2N	Otherwise Largest Control χ^2	Control Statistic Involved
2	44.44	92.42	88.68	75.00	123.36	40.25	6796.24	TY 3
4	27.71	15.40	40.58	17.50	30.86	30.43	39.59	ZY NN
8	13.37	37.42	42.27	38.22	30.58	25.16	37.67	ZY 2N2N
16	29.22	18.07	21.45	27.27	35.50	25.89	35.96	TY 2NN
32	20.85	23.45	20.61	18.72	30.86	40.85	44.62	TY 1234
64	21.77	15.25	23.55	28.91	20.89	18.11	32.30	TY 3NN
128	17.82	11.13	24.88	12.54	16.09	17.07	33.36	FY 2NNN
256	13.66	17.64	24.39	25.15	11.15	27.26	25.33	TY 3
512	32.15	27.16	24.94	13.76	14.15	25.64	46.25	FY NNNN
1024	21.63	17.26	29.78	22.07	22.12	15.18	26.24	TY 3

Symmetry Tests (11 degrees of freedom)

N	ZY 1	ZY 2	ZY 3	ZY 4	ZY 1234
2	12.13	9.93	9.90	14.13	15.88
4	4.34	4.20	12.61	5.67	7.53
8	3.36	12.88	11.70	8.47	4.67

The table shows that all of the "building block" statistics (ZY 1, ZY 2, ZY 3, and ZY 4) pass the symmetry test at $N \leq 8$ and the goodness-of-fit-to normality test at all N's above 8, even if the criterion of rejection be as gross as attainment of the .05 level of significance. On this basis alone sampling is accepted as "sufficiently random". However, having "accepted", it is of some interest to examine further evidence which, although not part of the criterion, nevertheless bears upon the validity of the conclusion that sufficient randomness obtained. There are 23 control distributions in all (i.e., distributions of statistics using observations from only the Y population) and Table VII shows that above $N = 2$ none of the chi-square tests on any of these distributions ever reached the .001 level of significance. Another question of interest is what would have happened if, instead of basing the tests upon the individual building blocks, an "overall" test embracing the entire pool of 4N random numbers had been conducted. One way of getting at this is simply to conduct the test of randomness upon ZY 1234. Table VII shows that except for attainment of the .05 level of significance at $N = 16$, this test was passed at all $N > 2$. Another way of getting an overall test is to sum the individual χ^2's for ZY 1, ZY 2, ZY 3, and ZY 4, thereby obtaining an "overall" χ^2 with $4(21) = 84$ degrees of freedom. This was done (but results are not shown in Table VII). At $N = 8$ the χ^2 was significant at the .001 level, but in all other cases above $N = 2$, χ^2 failed to fall in the .05 rejection region, thereby passing the test. Still another question of interest is whether or not a verdict of sufficient randomness would have been obtained had a test sensitive to the dispersion, rather than the means, of the pseudo-random numbers been used. Since RY 2N2N is the ratio of the variance of the observations selected by the first 2N pseudo-random numbers in the pool of 4N numbers to the variance of those selected by the second 2N, a chi-square test of goodness-of-fit to normal-theory upon RY 2N2N provides such a test (at least at $N > 8$). Table VII shows that, except at $N = 32$ where the .01 level of significance was attained, this test was passed at all $N > 2$. Finally, since means and variances of samples from a normal distribution are statistically independent, the ZY 1, ZY 2, ZY 3, and ZY 4 which involve means but not variances are (in effect, above $N = 8$) statistically independent of RY 2N2N which involves variances but not means. Thus, the sum of the χ^2's for the four ZY statistics and RY 2N2N yields a valid overall χ^2 with 5(21) = 105 degrees of freedom which is sensitive to both the means and variances of the pseudo-random numbers (although it is "weighted" so as to be, in a sense, four times as sensitive to means as to variances). This overall χ^2 was obtained at all values of N, and was found to fall in the .05 rejection region only at $N \leqq 8$, attaining only the .05 level of significance at N = 4, but reaching the .001 level at N = 8 and, of course, at N = 2. Thus all of the investigated statistical evidence tends to support the conclusion that sampling was quite "reasonably" random for all of the batches of data (N's) reported, but that at $N \leqq 8$ the discreteness of the Y population caused sampling distributions obtained from it to depart detectably from normal-theory and that these departures became quite appreciable at N = 2.

APPENDIX IV

Photographed Tables of Original Output Data

This appendix presents photographs of the output tables printed out by the machine. The virtue of photographs is that transcribing errors are thereby eliminated. For each statistic investigated, the tables give the empirical cumulative probabilities of those values of the statistic which under normal theory have the cumulative probabilities listed as column headings. Thus they give the actual empirical cumulative probabilities (under conditions of assumption violation) corresponding to cumulative probabilities which would be the true cumulative probabilities had all the assumptions been met. Hence the column headings list the alleged significance level at which normal-theory says a test is being conducted and the column entries list the empirical significance level, i.e., the empirical estimate of the true significance level, at which a test, listed as a row heading, is actually being conducted due to the violation of its normal-theory assumptions.

There are five sets of ten tables, each table in a given set corresponding to a different value of N, which is listed at the top of the table. The first set of ten tables pertains exclusively to the situation in which all of the observations upon which a multi-sample statistic is based were drawn from a single population, which is either the X population or the Y population. All of the statistics for which this was the case are listed in this set of tables and the identifying code lists the sampled population only once since all samples were drawn from a common population. This set of tables is completely redundant in that all of the statistics included in it appear later in one of the other sets of tables, none of which is redundant with regard to any set except the first one. There are four other sets. One set of ten tables presents all two-sample Z statistics. Another set of ten tables presents all two-sample T statistics. A third set of ten tables presents all F and R statistics. A final set of ten tables presents all one-sample Z and T statistics investigated in the present study plus the relevant data for the Z and T statistics investigated in the earlier study (and identified as ZX 0, ZY 0, TX 0, and TY 0 to indicate their earlier acquisition and their comparability with the statistics having the same code except that the zero is replaced by a 1, 2, 3 or 4), plus the corresponding relevant data for the sampling distributions obtained by combining what are, in effect, five independent sampling distributions of the same one-sample statistic. Thus the statistic-values comprising the independent sampling distributions of ZX 0, ZX 1, ZX 2, ZX 3, and ZX 4, each of which consists of 10,000 values of $Z = \frac{\bar{X} - 100}{\sqrt{\frac{125.70684}{N}}}$, are combined into a single sampling distribution consisting of 50,000 values of $Z = \frac{\bar{X} - 100}{\sqrt{\frac{125.70684}{N}}}$, which is now designated simply ZX. The statistics identified as ZY, TX, and TY, and their sampling distribution of 50,000 statistic-values each, are obtained analogously. The data for ZX 0, ZY 0, TX 0, TY 0, ZX, ZY, TX and TY were typed on to the IBM sheets by the writer, not by the computing machine. The entries for all statistics having empirical sampling distributions consisting of 10,000 values are simply empirical cumulative frequencies which can be presented in four or fewer digits (since cumulating only up to a value corresponding to a normal-theory α of .50 prevents the empirical cumulative frequency of 10,000 from occurring as an entry). These empirical cumulative frequencies when divided by 10,000 become empirical cumulative probabilities, and since the effect is simply to move the decimal four places to the left, the entries may be regarded as the significant digits of the empirical cumulative probabilities, i.e., as empirical cumulative probabilities with the decimal point and zeros preceding the first nonzero digit missing. However, for statistics having empirical sampling distributions based on 50,000 values the entries are empirical cumulative probabilities only. Consider the empirical cumulative probability of ZX corresponding to a given value of α, i.e., for a fixed column heading. This can be obtained by summing the empirical cumulative probabilities of ZX 0, ZX 1, ZX 2, ZX 3 and ZX 4 at the same value of α and multiplying the sum by .2. Thus, since a maximum of four digits is required to express the probabilities summed, a maximum of five digits is necessary to express the empirical cumulative probability of a statistic having an empirical sampling distribution based on 50,000 values. Therefore, the table entries for such statistics extend one place farther to the right than do those for statistics whose empirical sampling distributions contain only 10,000 values. In order to prevent confusion, digits the same distance from the decimal place are aligned in the entries for the two types of statistic.

A final table is concerned only with Z statistics and with one-tailed α's of .0001. It lists, for each Z statistic at each N value, the left-tailed empirical cumulative probability of a Z of -3.71901649 and the right-tailed empirical cumulative probability of a Z of +3.71901649. Thus it gives the empirical estimates of the true significance levels corresponding to a normal-theory significance level of .0001 for a left-tailed and for a right-tailed test, respectively. Again, the decimal point and the zeros preceding the first significant digit are omitted, so, except for ZX and ZY, the rightmost digit of an entry is the fourth figure to the right of the decimal, and for ZX and ZY it is the fifth figure to the right of the decimal. In the case of ZX and ZY, a period is inserted between the fourth and fifth figures to the right of the true decimal point. Entries for ZX 0, ZY 0, ZX and ZY were written in the table by the author. This table was not printed out as such by the machine, but was assembled from material printed out by the machine at the various values of N. The following material, also, was not printed out by the machine in the table in which it occurs, but was later cut out of the table in which it was originally printed out and spliced into the new table: ZX NN and ZY NN were inserted into the table devoted exclusively to Z statistics; TX NN and TY NN were inserted into the table devoted exclusively to T statistics; FX NONN, FY NONN, FX 2NNN, FY 2NNN, RX 2N2N and RY 2N2N were inserted into the table containing only F and R statistics.

As mentioned in the Results section, data for certain statistics, which, in effect, estimate essentially the same effect, were combined in obtaining the graphs presented in the Results section. Other statistics were not graphically dealt with at all in the Results section. They fall into two groups as follows: (a) the control statistics ZYY NN, ZYY 2NN, ZYY 3NN, ZYY 2N2N, TYY NN, TYY 2NN, TYY 3NN, TYY 2N2N, RY 2N2N, FYY NNON, FYY NONN, FYY 2NNN, FYY NNNN, ZY 1, ZY 2, ZY 3, ZY 4, ZY 1234, TY 1, TY 2, TY 3, TY 4, TY 1234, (b) the statistics FXX NONN, ZX 1234 and TX 1234 whose true sampling distributions were identical with those of FXX NNON (at the same N), ZX (at 4N), and TX (at 4N), respectively, but whose empirical sampling distributions were not combined with those of the corresponding statistic. Such combination was not attempted, even though it would presumably have improved precision of estimate, because it would have been obtained at the cost of introducing several new forms of nonindependence and imposed other qualifications upon the interpretation of the data while improving precision of estimate for only a very few statistics. (The ZX 1234 data at N = 512 and N = 1024 were used to extend ZX data to N = 2048 and N = 4096, and likewise for TX 1234 and TX, and the extended data was nonindependent, however, there was no combination of nonindependent data in these cases.)

N= 2

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX	NN	56	67	124	150	323	689	1197	1559	1653	2065	5428	5300	2104	1659	1565	1206	702	346	159	124	68	52	11600.22
ZY	NN	7	14	49	102	267	501	999	2093	3074	4115	5148	5017	4024	3038	2061	1060	542	265	96	50	6	3	35.29
ZX	2NN	93	106	172	246	508	871	1259	1579	1725	2121	4663	5676	3228	2678	1804	619	338	186	76	38	13	9	7234.17
ZY	2NN	6	11	58	99	256	497	996	2023	2960	4075	5095	5012	4042	2990	2015	1024	503	250	104	50	14	9	38.07
ZX	3NN	93	126	205	305	594	925	1290	1624	1758	2082	4015	6165	4118	2962	1373	522	195	77	31	11	2	2	6036.06
ZY	3NN	6	19	69	117	257	487	959	1936	3012	4067	5094	4993	4012	2994	1951	1008	503	261	107	56	17	11	46.20
ZX	2N2N	20	28	95	180	312	453	944	2072	2520	2873	5229	5155	2910	2531	2092	952	450	295	171	100	33	23	4200.19
ZY	2N2N	7	15	64	114	273	501	1018	1993	2920	3847	5050	5077	3922	2961	1979	1006	431	221	97	49	15	10	77.51
TX	NN	64	68	89	136	232	429	741	1749	3626	4237	5428	5300	4128	3532	1741	747	427	224	129	84	67	65	3412.54
TY	NN	8	14	45	95	256	503	968	2006	3058	4084	5148	5017	3941	2989	2011	997	496	236	103	49	7	6	21.66
TX	2NN	60	70	110	126	208	364	1477	2455	3147	3866	4663	5676	4926	3871	1518	585	276	148	54	34	7	3	3954.63
TY	2NN	7	13	56	95	250	474	983	1987	3017	4041	5095	5012	3974	2972	1992	992	514	250	97	49	7	5	12.61
TX	3NN	65	66	90	113	260	1010	1550	2224	2874	3428	4015	6165	5507	3529	1219	465	213	119	45	27	10	5	5640.43
TY	3NN	7	16	56	98	240	460	964	1968	2990	4035	5094	4993	3965	2954	1986	1008	501	269	103	58	13	9	21.60
TX	2N2N	1	3	10	26	118	228	723	2562	3663	4253	5229	5155	4259	3684	2598	739	259	117	43	16	4	2	2218.36
TY	2N2N	6	10	42	88	268	504	975	1971	2897	3892	5050	5077	3969	2932	1926	981	511	234	80	42	8	4	40.65
FX	NNON	85	85	87	90	148	288	565	1114	1730	2881	4540	5460	2953	1852	1332	736	442	295	194	139	71	57	5772.78
FY	NNON	3	8	52	95	230	456	997	1996	3047	4057	5078	4922	3953	2931	1957	1001	491	244	103	50	9	7	20.13
FX	NONN	78	78	81	85	137	289	554	1118	1704	2873	4514	5486	2981	1821	1303	716	425	276	180	131	68	50	5454.34
FY	NONN	7	12	55	110	251	[illegible]	1013	2020	3022	4017	5013	4987	3985	3005	1975	973	457	232	88	46	12	11	20.84
FX	2NNN	46	46	51	57	120	249	463	1060	2376	4013	5058	4942	4134	3087	1367	543	331	214	158	128	94	85	3529.32
FY	2NNN	4	12	60	112	279	544	1078	2068	3061	4086	5140	4860	3871	2903	2025	980	482	228	105	65	12	8	43.72
FX	NNNN	20	20	39	70	150	247	431	961	1602	2848	5078	4922	2575	1648	1137	667	430	322	243	194	117	93	6565.06
FY	NNNN	25	40	104	158	293	508	971	2039	3125	4128	5115	4885	3944	2957	1972	980	489	272	134	84	38	27	250.37
RX	2N2N	720	1105	2021	2245	2419	2544	2787	3296	3819	4389	5122	5095	4381	3810	3331	2816	2586	2465	2261	2057	1155	748	314767.02
RY	2N2N	4	9	58	109	246	482	1034	2029	2994	3994	4964	5052	4036	3086	2127	1085	523	254	93	52	17	5	40.25

N= 4

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX	NN	21	30	93	157	298	434	957	1991	2423	2833	5195	5169	2804	2396	1924	897	418	281	161	90	38	24	4377.65
ZY	NN	1	5	41	98	229	496	1047	2035	3038	3991	5082	5034	3927	3028	2090	1025	508	255	98	50	12	7	39.59
ZX	2NN	46	68	157	232	354	564	1103	1946	2481	2988	4581	5563	3999	3040	1625	748	302	155	66	37	10	6	1754.18
ZY	2NN	1	10	41	104	250	503	1012	2014	3052	4017	5044	5017	4028	3046	2021	1035	514	265	121	65	13	5	24.91
ZX	3NN	59	89	175	245	399	620	1108	1953	2554	3027	4106	5972	4584	3046	1641	563	199	84	29	16	2	1	2037.51
ZY	3NN	3	9	53	100	241	479	1013	2031	3001	3976	4966	5082	4034	3020	1987	1021	529	242	118	50	15	4	35.65
ZX	2N2N	17	22	67	119	267	514	940	1944	2941	3639	5087	5063	3661	2938	1921	933	516	272	125	80	30	18	515.55
ZY	2N2N	8	20	63	119	264	524	1025	2011	2993	4022	5086	5016	3949	2948	2018	1044	548	286	126	70	17	6	34.10
TX	NN	1	3	17	36	124	233	748	2599	3661	4245	5195	5169	4192	3601	2605	790	269	139	42	18	3	2	2086.30
TY	NN	6	11	42	92	253	491	1006	2006	3018	4061	5082	5034	3988	3035	2078	1041	515	247	110	48	9	4	19.87
TX	2NN	2	4	26	64	263	451	1439	2372	3136	3873	4581	5563	4945	4048	1978	414	142	62	22	14	4	4	2827.04
TY	2NN	3	8	54	114	251	478	986	2029	3049	4009	5044	5017	3997	3031	2076	1010	518	236	103	53	10	7	26.56
TX	3NN	5	9	109	161	336	872	1401	2166	2948	3558	4106	5972	5363	4054	1695	248	83	31	10	7	2	0	3820.12
TY	3NN	5	8	44	100	249	473	970	2008	2985	3987	4966	5082	4042	3014	2011	1019	486	248	83	39	9	6	18.24
TX	2N2N	0	1	9	16	72	250	941	2558	3528	4228	5087	5063	4212	3526	2575	983	264	61	21	8	2	1	1473.82
TY	2N2N	6	10	68	115	266	518	985	2012	3019	4037	5086	5016	3984	2958	2014	1051	508	259	119	59	10	5	25.92
FX	NNON	20	21	38	59	147	268	519	854	1397	2542	4025	5975	4064	2207	1314	520	187	97	39	16	4	3	3070.87
FY	NNON	3	3	42	90	238	481	979	1957	2928	3953	4948	5052	3999	3015	2009	1010	497	243	105	52	15	7	15.58
FX	NONN	22	22	32	52	156	288	489	838	1330	2459	3960	6040	4130	2204	1316	517	202	91	35	12	4	4	3289.95
FY	NONN	1	10	53	94	233	495	950	1943	2878	3869	4893	5107	4117	3049	2047	1055	506	253	98	45	14	9	33.81
FX	2NNN	8	8	16	42	115	218	398	918	1877	3349	4775	5225	4080	2823	1489	595	352	138	37	21	6	3	1414.52
FY	2NNN	9	11	52	90	243	478	967	1936	2950	3949	4973	5027	4043	3008	2008	1012	525	268	106	49	6	5	18.24
FX	NNNN	2	4	31	55	106	172	285	700	1476	2714	4583	5417	3484	2123	1242	617	178	93	48	28	4	4	3100.89
FY	NNNN	3	7	53	97	243	480	999	1933	2981	3982	5012	4988	4023	3039	2066	1047	522	272	117	53	14	7	21.22
RX	2N2N	2358	2417	2524	2574	2686	2828	3110	3601	4075	4558	5008	5012	4555	4072	3599	3078	2841	2685	2578	2529	2429	2369	2228945.63
RY	2N2N	0	7	44	91	248	515	1042	2051	2991	3993	4988	5015	4065	3053	2082	1067	538	260	83	42	6	2	30.43

N= 8

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX	NN	24	29	87	134	273	524	933	1885	2896	3587	5005	5134	3686	2942	1891	927	492	253	117	62	18	10	547.85
ZY	NN	4	8	46	98	267	517	1035	2020	2985	3974	4995	5098	4019	3013	2034	999	488	235	92	45	11	7	11.30
ZX	2NN	40	51	136	197	364	622	1046	1894	2817	3620	4614	5422	4333	3094	1885	795	339	142	55	22	7	3	564.96
ZY	2NN	3	12	48	110	265	501	977	1949	3016	3999	5001	5044	3993	3026	1998	1026	505	231	93	49	9	5	23.90
ZX	3NN	49	63	159	220	398	651	1089	1925	2787	3633	4492	5540	4475	3276	1978	707	223	74	24	10	0	0	962.13
ZY	3NN	7	12	56	104	267	502	1008	1986	2998	4047	5069	4958	4010	3012	2030	995	521	241	103	54	9	5	18.91
ZX	2N2N	13	23	62	105	254	512	974	1909	2906	3843	4946	5119	3976	2973	1932	943	483	261	120	68	12	7	63.31
ZY	2N2N	9	18	58	108	245	483	964	1927	2921	3905	5013	5052	3954	2998	2059	1009	485	255	93	39	7	1	37.67
TX	NN	0	3	13	26	79	267	994	2525	3468	4146	5005	5134	4280	3575	2551	953	277	80	24	9	3	2	1352.18
TY	NN	3	9	59	113	245	487	1011	1968	2965	3972	4995	5098	4037	3014	2008	974	450	239	88	46	12	5	22.00
TX	2NN	3	9	52	100	339	663	1314	2224	3046	3830	4614	5422	4661	3765	2361	600	97	20	6	5	2	1	1413.70
TY	2NN	1	7	52	99	246	516	985	1957	3004	3969	5001	5044	3986	3010	1986	1035	496	220	87	48	5	3	29.83
TX	3NN	17	32	121	223	415	755	1269	2116	2933	3732	4492	5540	4707	3829	2401	385	30	6	3	2	1	1	2118.76
TY	3NN	7	11	47	100	242	500	990	1995	2996	4031	5069	4958	3989	3004	1990	1000	501	253	92	44	8	4	8.52
TX	2N2N	0	1	9	28	145	452	1118	2201	3198	4021	4946	5119	4161	3283	2257	1057	418	143	25	6	0	0	345.37
TY	2N2N	5	12	51	94	250	482	969	1935	2922	3924	5013	5052	3973	3035	2073	1002	489	251	91	37	8	3	26.67
FX	NNON	6	8	37	74	184	338	620	1144	1829	2886	4194	5806	4336	2731	1590	595	234	85	37	15	3	2	1370.27
FY	NNON	6	7	50	110	269	514	1033	2030	3079	4049	5021	4979	3995	3019	1980	762	477	271	111	44	7	3	29.62
FX	NONN	3	4	36	69	179	347	623	1113	1804	2868	4183	5817	4355	2798	1644	605	257	93	32	16	4	3	1334.04
FY	NONN	5	12	54	106	256	501	1009	1992	2990	3948	4977	5023	4022	3018	1972	999	518	266	104	51	13	6	9.95
FX	2NNN	3	5	33	68	162	324	616	1264	2114	3266	4607	5393	4037	2817	1674	765	384	185	89	45	11	6	698.75
FY	2NNN	2	7	54	103	253	519	1064	2055	3022	4032	5002	4998	4004	3009	2032	1031	519	253	98	54	9	5	13.90
FX	NNNN	3	5	30	60	150	262	459	1086	2064	3309	4713	5287	3899	2583	1576	734	306	156	55	30	5	4	984.39
FY	NNNN	3	6	52	97	275	535	1050	2028	3033	3997	4999	5001	3995	3005	1989	1021	541	277	126	59	14	8	22.30
RX	2N2N	1666	1720	1939	2092	2374	2687	3061	3661	4139	4571	4981	5020	4630	4195	3739	3161	2728	2405	2131	1979	1733	1671	1112873.92
RY	2N2N	1	3	37	98	244	515	1038	2052	3029	3966	4970	5032	4049	3055	2054	1019	524	259	119	52	11	5	25.16

N= 16

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES												RIGHT-TAIL CUMULATIVE PROBABILITIES												
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX	NN	9	23	85	142	280	514	991	1954	2948	3937	5056	5007	3904	2920	1912	965	489	288	128	70	19	8	77.67
ZY	NN	5	10	43	97	250	500	1031	2061	2980	3982	5076	5000	3910	2920	1942	980	470	225	94	43	9	5	22.69
ZX	2NN	30	42	116	182	365	629	1056	1946	2840	3808	4805	5229	4196	3090	1936	876	358	175	64	21	1	0	299.01
ZY	2NN	5	9	47	86	255	515	1005	2004	3035	4057	5112	4928	3928	2922	1950	969	443	215	80	37	10	5	19.97
ZX	3NN	34	52	133	198	376	622	1071	1933	2791	3728	4712	5309	4312	3145	1986	842	328	122	35	10	1	1	455.54
ZY	3NN	7	11	50	102	251	494	1001	2012	3043	4089	5050	4978	3905	2909	1954	974	462	215	73	41	11	5	23.27
ZX	2N2N	4	8	61	111	265	516	997	2041	3010	4042	5121	4936	3874	2907	1979	983	459	253	107	50	8	2	31.42
ZY	2N2N	4	10	43	90	235	492	1014	2015	2973	4010	5065	4976	3917	2894	1925	940	445	214	88	49	9	4	18.61
TX	NN	1	1	8	32	170	472	1124	2274	3261	4119	5056	5007	4079	3235	2294	1106	449	154	31	5	0	0	346.97
TY	NN	4	10	47	98	254	506	1017	2040	2991	4007	5076	5000	3941	2930	1966	984	480	225	73	36	11	6	16.84
TX	2NN	10	16	80	160	364	670	1203	2120	2997	3889	4805	5229	4275	3313	2215	971	260	46	2	0	0	0	455.41
TY	2NN	5	7	44	87	237	543	1005	1987	3014	4052	5112	4928	3918	2940	1960	961	440	199	75	40	9	4	35.96
TX	3NN	24	38	125	200	400	658	1135	2053	2893	3793	4712	5309	4370	3315	2203	947	187	27	2	0	0	0	678.87
TY	3NN	5	9	44	91	257	490	1007	2010	3035	4089	5050	4978	3908	2935	1987	979	458	212	80	39	6	2	25.39
TX	2N2N	0	0	24	72	221	531	1116	2170	3142	4082	5121	4936	3941	3036	2102	1047	481	216	67	21	0	0	99.24
TY	2N2N	0	9	40	91	237	486	1019	2003	2991	4001	5065	4976	3899	2920	1947	933	436	209	85	42	8	5	28.80
FX	NNON	5	8	44	99	242	439	830	1690	2564	3514	4492	5508	4413	3218	1950	821	343	145	53	28	3	2	268.34
FY	NNON	3	10	56	107	244	495	1027	2016	2982	3979	4976	5024	4020	3016	2025	979	480	221	92	46	8	5	13.71
FX	NONN	4	10	40	83	230	420	813	1653	2581	3551	4525	5475	4300	3135	1930	796	355	164	57	26	3	2	244.09
FY	NONN	3	10	52	108	248	497	1007	2051	3062	4055	5109	4891	3969	2973	1965	993	490	246	91	37	9	7	22.11
FX	2NNN	3	7	45	102	226	457	873	1776	2674	3674	4716	5284	4224	3056	1890	859	429	234	78	46	13	11	142.91
FY	2NNN	6	14	45	97	239	512	1027	2033	3037	4002	5025	4975	3951	2932	1924	944	443	210	87	37	6	3	19.95
FX	NNNN	5	12	57	90	206	408	817	1673	2615	3601	4660	5340	4171	3025	1834	865	407	175	67	41	13	9	190.89
FY	NNNN	7	11	48	101	247	501	1025	2010	3028	4022	4987	5013	3987	2991	1978	976	467	240	90	46	8	1	12.50
RX	2N2N	1184	1307	1713	1934	2290	2652	3116	3725	4231	4667	5044	4956	4554	4103	3627	3045	2632	2264	1899	1661	1279	1147	555342.40
RY	2N2N	6	6	41	98	244	500	1074	2047	3069	4080	5028	4973	3966	2975	2008	1059	547	286	110	65	9	5	25.89

N= 32

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES										RIGHT-TAIL CUMULATIVE PROBABILITIES												
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX	NN	6	9	47	97	233	483	983	1934	2913	3961	5021	5017	3998	2939	1943	982	471	242	97	50	10	8	20.20
ZY	NN	6	11	42	86	242	469	943	1955	2968	3953	5028	5034	4007	2983	2030	1049	520	257	111	54	8	5	17.43
ZX	2NN	8	13	66	115	285	550	1022	1904	2828	3817	4834	5191	4175	3065	1958	894	405	198	71	34	8	5	78.92
ZY	2NN	3	8	43	81	223	472	978	1981	2937	3949	4982	5041	4016	2993	1995	1011	512	265	105	56	15	10	14.08
ZX	3NN	13	20	74	146	318	578	1067	1938	2842	3764	4720	5300	4275	3134	2004	878	383	174	43	17	4	1	162.71
ZY	3NN	6	10	40	90	237	477	987	1947	2958	3972	4977	5044	4062	2992	2027	1024	526	288	118	58	12	7	19.40
ZX	2N2N	5	9	44	91	223	452	947	1951	2910	3957	4973	5056	4013	2987	1964	939	464	239	105	58	20	12	28.87
ZY	2N2N	5	5	42	93	230	454	979	1998	2977	3954	4977	5064	3939	2993	2020	988	498	259	106	55	15	7	30.62
TX	NN	0	0	18	58	201	478	1056	2060	3071	4026	5021	5017	4071	3101	2083	1048	491	207	76	24	1	1	79.06
TY	NN	2	9	45	100	249	467	950	1949	2999	3945	5028	5034	3991	2994	2017	1047	524	261	112	50	5	3	25.73
TX	2NN	2	8	53	115	291	555	1041	1971	2862	3862	4834	5191	4224	3163	2063	977	393	170	36	6	0	0	118.43
TY	2NN	5	9	43	91	226	474	996	1961	2936	3944	4982	5041	4029	3004	1994	1014	513	259	103	57	13	10	14.23
TX	3NN	3	12	77	144	322	595	1080	1979	2882	3794	4720	5300	4284	3220	2105	938	383	128	21	4	0	0	195.51
TY	3NN	6	8	43	89	240	482	1005	1953	2981	3973	4977	5044	4066	3006	2003	1017	512	271	113	51	12	6	16.83
TX	2N2N	0	2	34	81	222	492	978	1990	2982	3996	4973	5056	4037	3051	2005	991	496	238	93	46	6	0	21.30
TY	2N2N	5	6	48	91	240	450	977	1987	2963	3962	4977	5064	3950	2998	2003	983	493	262	108	55	12	6	26.88
FX	NNON	5	9	46	89	217	419	912	1904	2860	3849	4839	5161	4176	3073	1962	915	394	168	55	21	4	3	76.19
FY	NNON	5	9	59	112	258	487	1031	2031	3052	4022	5029	4971	3989	3036	2052	1007	493	260	104	62	10	5	21.49
FX	NONN	3	7	46	82	220	444	909	1787	2779	3797	4795	5205	4148	3071	2025	920	427	168	55	21	4	3	80.35
FY	NONN	6	13	57	111	259	501	1004	1990	2975	4013	5009	4991	4030	3015	2026	1006	506	262	104	47	10	4	8.34
FX	2NNN	2	8	52	102	233	470	923	1913	2890	3914	4936	5064	4062	3046	2010	951	423	207	75	24	[illegible]	2	38.60
FY	2NNN	5	11	49	101	239	523	986	2001	3019	4015	5006	4994	4036	3040	2032	1011	515	246	98	50	12	5	14.12
FX	NNNN	8	15	52	89	228	455	920	1866	2807	3782	4831	5169	4115	2988	1944	913	407	189	53	26	4	2	70.70
FY	NNNN	5	11	50	106	249	514	1012	2008	3009	3967	4950	5050	4035	3019	2027	1020	522	253	116	58	10	3	11.90
RX	2N2N	1003	1126	1539	1790	2204	2595	3086	3704	4165	4558	4971	5029	4657	4243	3713	3065	2592	2190	1770	1547	1122	987	410177.08
RY	2N2N	3	6	67	116	290	514	1005	2041	2993	3965	5025	4975	3922	2988	1980	980	485	222	93	53	10	4	40.85

N= 64

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES										RIGHT-TAIL CUMULATIVE PROBABILITIES												
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX	NN	7	10	59	106	270	526	1009	2036	3051	4025	5071	4951	3962	2935	1977	961	494	249	112	56	5	3	21.67
ZY	NN	4	12	58	104	237	493	994	2016	2945	3980	4986	5046	3982	2971	1960	994	481	239	86	42	8	4	18.52
ZX	2NN	9	13	76	148	322	587	1064	2036	2984	3946	4958	5061	4037	3000	1931	926	417	197	70	28	4	2	61.98
ZY	2NN	11	13	56	111	289	530	1035	1996	2969	3947	4975	5039	4023	2992	1941	959	457	220	91	43	4	1	30.20
ZX	3NN	10	26	86	162	354	609	1109	2012	301[illegible]	3941	4917	5093	4124	3114	1994	944	404	165	55	24	4	0	135.08
ZY	3NN	8	15	65	118	261	533	1037	2032	297[illegible]	3951	4989	5024	4005	2997	1956	975	477	238	98	35	9	2	26.49
ZX	2N2N	6	10	39	85	241	508	1027	2057	3085	4088	5093	4929	3915	2925	1936	938	448	210	91	55	9	3	20.83
ZY	2N2N	4	9	47	107	255	492	995	2028	3010	3985	5000	5020	4017	2975	1969	967	469	235	92	44	7	5	10.40
TX	NN	3	6	49	89	238	513	1023	2062	3090	4057	5071	4951	4001	2983	2011	1016	490	240	91	42	5	1	19.79
TY	NN	7	13	53	103	246	491	995	1990	2956	3995	4986	5046	3990	2983	1964	983	491	240	92	40	8	4	9.40
TX	2NN	7	11	68	132	314	565	1098	2061	3017	3966	4958	5061	4047	3063	1989	956	432	193	46	21	1	0	66.60
TY	2NN	6	14	58	111	277	529	1042	1995	2962	3949	4975	5039	4024	2998	1945	962	460	235	88	42	7	2	17.62
TX	3NN	12	17	74	159	347	623	1108	2063	3014	3964	4917	5093	4147	3135	2037	957	405	170	40	14	0	0	130.96
TY	3NN	10	12	66	119	267	532	1030	2043	2978	3946	4989	5024	4014	2997	1968	974	469	229	101	43	11	2	32.30
TX	2N2N	2	8	33	80	223	505	1027	2088	3131	4089	5093	4929	3952	2965	1977	952	460	228	84	37	4	1	29.74
TY	2N2N	5	9	51	105	254	504	1006	2005	3028	3990	5000	5020	4014	3002	1961	960	479	233	96	40	11	4	11.11
FX	NNON	3	8	48	98	244	482	977	1898	2875	3868	4913	5087	4071	3047	1991	960	476	230	78	39	7	5	21.73
FY	NNON	6	14	61	115	253	509	1032	1989	3007	4030	5017	4983	3941	2940	1958	966	488	230	99	54	13	4	18.29
FX	NONN	6	.10	42	87	231	471	960	1917	2876	3897	4939	5061	4043	3042	1976	951	471	220	83	38	6	2	20.45
FY	NONN	5	9	54	104	256	493	1027	2015	2989	3999	4979	5021	3968	2944	1936	989	502	280	109	44	14	8	26.83
FX	2NNN	4	10	53	87	227	451	940	1904	2901	3910	4906	5094	4088	3014	2008	966	480	219	89	43	15	8	28.13
FY	2NNN	2	6	46	100	254	496	1005	1995	3021	4036	5033	4967	4003	2998	1982	993	540	276	119	55	11	7	16.40
FX	NNNN	4	5	38	84	247	464	913	1873	2864	3893	4930	5070	4019	2976	1933	987	466	241	90	52	11	5	35.23
FY	NNNN	4	5	47	90	268	507	995	1980	3000	4006	5050	4950	3955	2971	1978	966	497	269	106	52	10	6	19.35
RX	2N2N	1039	1194	1619	1889	2311	2707	3175	3819	4270	4702	5096	4904	4494	4080	3623	2996	2524	2085	1685	1444	1052	911	398083.55
RY	2N2N	5	7	48	101	253	496	1021	2040	3080	4090	5107	4893	3898	2897	1913	979	512	239	88	42	11	7	18.11

N= 128

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX	NN	4	10	52	102	241	477	980	1951	2952	3937	4933	5084	4089	3099	2041	994	502	250	94	38	11	8	17.12
ZY	NN	7	10	50	91	241	488	970	1966	2946	3942	4917	5105	4047	3048	2060	1022	513	244	78	44	9	4	18.69
ZX	2NN	10	15	74	136	270	488	948	1877	2829	3830	4807	5202	4153	3127	2065	964	470	225	72	30	4	2	60.72
ZY	2NN	3	10	55	109	240	482	953	1912	2956	3937	4954	5055	4050	3080	2062	1041	494	247	87	42	7	3	19.93
ZX	3NN	11	22	81	134	267	499	953	1891	2810	3778	4778	5229	4191	3107	2092	961	439	195	64	25	1	0	88.67
ZY	3NN	1	7	52	104	239	474	964	1955	2976	3958	4896	5110	4064	3080	2106	1017	503	226	87	45	5	2	30.47
ZX	2N2N	4	9	55	99	263	506	977	1927	2921	3945	4924	5086	4039	3026	2046	1044	513	247	79	33	9	4	23.16
ZY	2N2N	4	7	46	99	229	496	988	1921	2948	3947	4971	5049	4009	3031	2005	993	488	233	87	46	8	3	15.55
TX	NN	2	5	43	95	222	484	982	1974	2972	3967	4933	5084	4131	3114	2062	1009	497	242	95	44	8	2	20.54
TY	NN	6	10	51	92	232	488	954	1939	2955	3946	4917	5105	4044	3040	2053	1035	516	241	81	45	7	4	19.25
TX	2NN	10	17	65	117	262	486	987	1905	2828	3826	4807	5202	4153	3122	2073	991	453	213	69	26	2	0	53.34
TY	2NN	3	7	55	106	236	475	956	1914	2940	3928	4954	5055	4045	3072	2048	1040	497	239	85	45	7	3	18.10
TX	3NN	8	17	83	135	258	507	972	1903	2820	3782	4778	5229	4202	3101	2089	964	444	191	63	24	0	0	91.85
TY	3NN	1	8	50	104	244	479	970	1958	2972	3962	4896	5110	4067	3087	2106	1019	495	230	90	44	5	0	29.34
TX	2N2N	6	13	46	100	253	502	1000	1935	2942	3943	4924	5086	4045	3049	2043	1056	516	238	75	31	7	4	24.57
TY	2N2N	4	8	42	98	236	500	984	1921	2948	3949	4971	5049	4005	3046	2009	996	489	237	86	45	8	3	16.26
FX	NNON	0	9	46	105	260	510	947	1992	2974	4010	4986	5014	3965	2967	1981	956	436	191	86	46	15	9	47.42
FY	NNON	5	6	41	103	251	499	973	2004	2974	3988	4946	5054	4001	3011	1993	1008	479	257	87	39	8	1	28.97
FX	NONN	2	8	44	82	230	471	988	1935	2990	3984	5003	4997	3991	3003	1999	976	473	214	92	46	3	2	24.73
FY	NONN	8	14	55	98	255	501	933	1986	3038	4022	4989	5011	4045	3042	1982	955	462	230	90	38	6	1	31.94
FX	2NNN	6	15	49	101	243	513	1046	2022	3015	4020	5063	4937	3987	3009	1961	980	497	247	109	52	9	4	19.68
FY	2NNN	8	9	32	84	232	500	953	1945	2991	4022	5060	4940	3991	2951	1973	971	479	244	93	36	7	4	33.36
FX	NNNN	6	10	50	96	245	483	1000	2001	3018	4036	5009	4991	4002	3013	1970	933	448	232	95	47	9	6	14.06
FY	NNNN	2	8	51	115	248	479	979	1990	2975	3968	4953	5047	4014	3036	1978	999	484	233	85	40	5	2	19.72
RX	2N2N	966	1113	1514	1753	2141	2525	3016	3684	4126	4538	4921	5079	4685	4283	3730	3103	2616	2206	1811	1579	1143	1013	406794.43
RY	2N2N	6	10	47	95	239	485	993	1946	2930	3961	4969	5031	3980	2997	2016	1015	471	240	83	39	8	3	17.07

N= 256

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX	NN	5	8	60	103	254	513	1013	1999	2938	3944	4943	5070	4009	2946	1942	971	453	218	93	43	10	7	26.06
ZY	NN	4	8	48	107	246	505	992	2043	3018	4049	5032	4984	3985	2999	1975	926	486	230	96	48	6	2	22.79
ZX	2NN	10	16	60	116	267	535	1046	1982	2945	3921	4916	5088	4048	3072	2050	958	449	214	77	40	10	3	34.59
ZY	2NN	3	9	46	102	234	480	1009	2005	3005	4010	5048	4960	3995	2953	1954	986	490	254	103	42	4	3	18.70
ZX	3NN	14	22	73	130	285	531	1023	1970	2945	3908	4897	5111	4073	3063	1991	964	459	208	72	42	6	3	45.92
ZY	3NN	4	7	55	102	255	481	997	2013	2990	4036	5063	4943	3940	2959	1986	959	489	256	101	43	5	3	18.05
ZX	2N2N	4	9	44	89	265	534	1018	1991	2928	4015	5030	4983	4013	3014	2039	992	512	248	84	37	2	1	34.39
ZY	2N2N	7	14	51	97	258	487	979	1984	3011	4047	5051	4961	4036	3014	2031	1014	506	239	96	46	7	4	17.16
TX	NN	2	7	52	104	256	504	1018	1990	2967	3956	4943	5070	4026	2972	1949	980	452	204	83	42	9	4	20.80
TY	NN	5	7	52	107	248	501	1002	2041	3019	4053	5032	4984	3990	3008	1969	931	483	237	90	50	7	4	19.75
TX	2NN	9	13	63	115	273	530	1051	1978	2954	3944	4916	5088	4050	3074	2065	955	447	199	81	38	5	2	40.22
TY	2NN	3	11	46	104	235	490	997	1995	3000	4013	5048	4960	3998	2959	1945	979	498	254	98	42	6	4	17.13
TX	3NN	13	23	79	126	288	537	1000	1974	2937	3912	4897	5111	4080	3067	1987	968	464	201	67	35	5	2	52.77
TY	3NN	4	5	57	106	252	482	992	2023	2987	4037	5063	4943	3939	2956	1983	957	487	261	90	44	7	3	24.50
TX	2N2N	4	5	35	88	267	519	1042	1973	2942	4012	5030	4983	4013	3017	2038	987	506	239	86	37	2	1	36.89
TY	2N2N	8	14	50	103	261	479	975	1989	3024	4046	5051	4961	4032	3013	2034	1008	510	239	96	48	8	4	19.03
FX	NNON	7	8	46	100	252	514	1029	2025	2999	4012	5039	4961	3995	3013	2013	991	483	242	89	42	5	3	12.54
FY	NNON	4	8	44	96	240	482	974	1999	3041	4066	5050	4950	3963	3020	2018	1007	486	244	99	46	14	11	19.40
FX	NONN	4	10	56	108	240	520	1037	2041	3075	4055	5049	4951	3992	3021	2010	993	480	239	101	42	7	2	17.92
FY	NONN	2	5	47	90	230	480	987	2001	3036	4023	5030	4970	3996	3012	2046	997	488	243	103	47	8	6	14.25
FX	2NNN	7	13	57	103	237	494	1004	1964	3013	4030	5001	4999	4000	3002	2037	995	504	234	88	40	7	3	16.12
FY	2NNN	4	6	38	89	216	467	1001	1979	2998	3996	5031	4969	3995	2995	2025	987	490	260	103	49	10	6	17.25
FX	NNNN	3	9	53	109	276	524	1018	2044	3035	4026	4984	5016	4030	3051	2043	987	484	237	85	42	5	2	14.46
FY	NNNN	3	8	44	84	244	474	976	2007	3030	4032	5025	4975	4043	3047	2027	1007	501	245	101	53	14	5	16.22
RX	2N2N	1031	1192	1569	1822	2192	2548	3047	3650	4153	4588	4958	5042	4652	4248	3770	3094	2612	2238	1834	1560	1108	969	416848.70
RY	2N2N	5	13	67	134	292	548	1052	1976	2965	3988	4958	5043	4007	3026	2008	979	475	243	90	49	12	5	27.26

N= 512

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX	NN	4	10	52	107	267	500	1047	2042	3040	4020	5034	4976	3979	2972	2024	1001	517	252	108	51	10	7	15.35
ZY	NN	4	10	52	97	222	496	1012	1994	3003	3962	4965	5045	4040	3005	1987	978	467	224	84	44	4	1	18.81
ZX	2NN	5	12	71	126	298	530	1036	2029	2966	3948	4942	5062	4032	3054	2006	1003	508	262	109	43	7	4	29.60
ZY	2NN	7	8	45	92	251	495	976	1919	2942	3919	4937	5074	4036	3003	1960	984	468	237	91	41	8	5	18.93
ZX	3NN	9	15	69	117	301	561	1030	1984	2951	3928	4929	5074	4103	3028	2070	1030	503	273	98	43	4	2	43.41
ZY	3NN	4	8	41	83	237	477	969	1929	2869	3900	4955	5049	4028	3010	1966	983	494	234	93	39	5	3	20.42
ZX	2N2N	2	11	53	103	249	512	1010	2022	2994	3987	4980	5026	3948	2932	1960	989	494	247	101	55	14	7	16.61
ZY	2N2N	5	11	52	120	275	505	1012	1979	2977	3959	4994	5012	4000	3011	1953	977	466	246	114	46	9	5	28.06
TX	NN	3	8	43	95	254	495	1044	2035	3047	4013	5034	4976	3971	2981	2027	1017	522	256	108	54	9	8	14.17
TY	NN	5	9	49	98	221	492	1004	1995	3004	3960	4965	5045	4035	3010	1981	974	471	223	80	44	4	1	19.05
TX	2NN	5	10	67	117	294	550	1028	2074	2953	3946	4942	5062	4042	3059	2018	1020	513	266	108	43	9	4	26.80
TY	2NN	5	9	44	93	247	497	975	1921	2938	3921	4937	5074	4037	2998	1971	983	470	237	90	40	8	5	12.86
TX	3NN	8	19	73	122	301	566	1035	1993	2953	3928	4929	5074	4101	3023	2075	1053	511	272	100	42	6	2	47.34
TY	3NN	3	8	42	84	241	475	961	1924	2870	3904	4955	5049	4029	3012	1961	990	494	235	89	39	5	3	20.58
TX	2N2N	1	10	49	101	254	507	1010	2043	3001	3989	4980	5026	3962	2942	1966	1000	498	243	105	61	16	7	22.03
TY	2N2N	4	10	53	121	273	508	1011	1984	2975	3962	4994	5012	3996	3012	1963	984	473	252	113	47	9	6	23.77
FX	NNON	3	8	38	107	240	519	990	2001	2991	3917	4931	5069	4042	3037	2057	1019	505	268	103	54	12	7	29.47
FY	NNON	4	8	62	104	257	478	1026	2047	3116	4100	5098	4902	3877	2901	1948	947	467	226	93	45	11	10	35.72
FX	NONN	7	13	51	102	267	495	999	1938	2961	3963	4936	5064	4044	3059	2060	987	516	268	92	50	10	6	23.41
FY	NONN	4	9	51	97	245	487	1012	2005	3018	4015	5015	4985	3993	2992	2013	979	472	231	81	34	8	3	10.34
FX	2NNN	3	4	40	89	245	510	999	2027	3006	3977	4992	5008	4052	3043	2024	1059	541	256	106	52	10	7	19.26
FY	2NNN	7	10	65	116	273	520	1059	2073	3089	4117	5069	4931	3978	2983	1985	1016	493	215	94	43	8	3	26.63
FX	NNNN	2	7	57	99	245	503	1013	2017	2967	3945	4949	5051	4045	3033	2041	1009	521	258	102	50	16	7	16.61
FY	NNNN	6	8	45	90	251	512	1061	2070	3095	4094	5082	4916	3920	3004	2013	957	432	238	89	37	12	10	46.25
RX	2N2N	950	1105	1562	1773	2155	2554	3052	3687	4155	4611	4997	5003	4613	4196	3699	3034	2532	2144	1771	1513	1064	904	362484.89
RY	2N2N	4	10	49	93	228	485	1006	2007	2950	3970	5053	4948	3978	2957	1926	977	516	263	96	51	8	5	25.64

N=1024

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX	NN	3	10	51	92	239	479	981	1949	2937	3944	4947	5063	4019	3013	2043	1028	519	246	108	52	10	5	11.67
ZY	NN	4	8	55	99	243	473	980	1981	2992	3978	4962	5046	4032	3030	2023	1004	526	253	114	57	13	6	11.79
ZX	2NN	4	11	57	103	265	500	984	1961	2939	3927	4922	5083	4084	3031	2013	995	475	240	84	50	16	8	20.83
ZY	2NN	4	12	49	96	248	488	962	1978	3008	4039	5006	4997	4000	3013	2035	1000	512	225	100	55	13	7	20.92
ZX	3NN	5	9	61	111	265	505	992	1970	2893	3904	4924	5079	4051	3009	2026	997	489	226	96	48	9	5	18.89
ZY	3NN	5	11	48	89	224	480	993	1992	2968	4008	5004	4998	4007	3030	2051	1009	502	258	99	44	12	7	12.74
ZX	2N2N	3	4	47	99	252	486	1000	2011	2996	3978	4953	5054	4028	2978	1990	1054	544	277	107	49	7	3	20.91
ZY	2N2N	5	9	49	100	238	489	967	1950	2946	4018	5024	4983	3973	2961	1950	1016	521	276	98	49	8	2	19.73
TX	NN	3	7	47	93	238	480	995	1957	2934	3934	4947	5063	4022	3017	2031	1032	508	238	109	58	11	3	15.28
TY	NN	4	9	55	98	244	477	989	1986	2994	3979	4962	5046	4032	3027	2028	1011	522	251	116	58	15	5	14.80
TX	2NN	5	13	60	103	257	504	995	1963	2942	3916	4922	5083	4089	3034	2010	993	479	245	89	49	15	6	18.38
TY	2NN	6	11	50	98	249	492	962	1975	3001	4037	5006	4997	4002	3016	2033	1002	513	224	102	55	12	7	19.75
TX	3NN	6	9	58	113	266	505	999	1967	2896	3917	4924	5079	4053	3007	2037	999	488	221	91	47	9	4	20.82
TY	3NN	6	13	44	90	225	486	992	1996	2971	4007	5004	4998	4005	3029	2053	1010	503	255	97	46	11	7	12.60
TX	2N2N	3	5	54	101	257	484	997	2009	2994	3984	4953	5054	4038	2982	1991	1046	541	277	107	53	7	1	23.68
TY	2N2N	4	10	45	103	237	494	976	1953	2950	4016	5024	4983	3970	2962	1957	1008	521	275	98	49	8	3	18.66
FX	NNON	4	5	53	98	258	527	1011	2017	2996	4047	5055	4945	3963	2981	1952	973	482	243	97	43	13	5	18.57
FY	NNON	5	13	50	92	253	489	965	1950	2975	4028	5060	4940	3912	2944	1950	980	499	256	101	55	9	3	16.78
FX	NONN	4	10	59	123	276	538	1036	2036	3002	3974	4969	5031	4016	3014	1990	978	501	263	94	30	6	3	25.86
FY	NONN	4	10	38	92	251	482	954	1983	3007	3991	4999	5001	3964	2976	1999	985	478	229	92	41	12	4	18.12
FX	2NNN	0	4	49	108	233	488	1024	2013	3020	4011	5014	4986	4002	3022	2040	989	500	265	92	37	10	2	31.14
FY	2NNN	4	8	45	101	246	477	988	1985	3018	4038	5013	4987	3971	2981	2034	999	501	231	86	43	7	1	16.27
FX	NNNN	3	8	57	115	275	524	1012	2011	3020	4009	4988	5012	3968	2983	1975	999	497	252	81	42	6	2	16.38
FY	NNNN	1	6	39	84	217	451	943	2004	3017	4061	5057	4943	3977	2994	2003	1005	470	216	97	46	7	2	26.14
XX	2N2N	914	1056	1483	1703	2149	2547	3052	3681	4140	4551	4956	5044	4642	4205	3730	3103	2587	2214	1794	1549	1095	960	367602.01
RY	2N2N	2	4	43	100	243	501	995	2015	2976	3968	4926	5075	4091	3093	2066	1013	512	251	93	46	9	3	15.18

N= 2

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZYY 2NN	6	11	58	99	256	497	996	2023	2960	4075	5095	5012	4042	2990	2015	1024	503	250	104	50	14	9	38.07
ZYY 3NN	6	19	69	117	257	487	959	1936	3012	4067	5094	4993	4012	2994	1951	1008	503	261	107	56	17	11	46.20
ZYY 2N2N	7	15	64	114	273	501	1018	1993	2920	3847	5050	5077	3922	2961	1979	1006	431	221	97	49	15	10	77.51
ZXX 2NN	93	106	172	246	508	871	1259	1579	1725	2121	4663	5676	3228	2678	1804	619	338	186	76	38	13	9	7234.17
ZXX 3NN	98	126	205	305	594	925	1290	1624	1758	2082	4015	6165	4118	2962	1373	522	195	77	31	11	2	2	6036.06
ZXX 2N2N	20	28	95	180	312	453	944	2072	2520	2873	5229	5155	2910	2531	2092	952	450	295	171	100	33	23	4200.19
ZXY 2NN	0	0	23	59	175	395	911	2035	3119	4196	5243	4862	3874	2867	1933	1049	586	340	186	105	33	25	212.90
ZXY 3NN	1	4	41	70	214	434	963	2014	3050	4098	5133	4934	3959	2967	1942	997	552	292	149	84	24	17	82.78
ZXY 2N2N	0	0	8	23	112	322	849	1959	3147	4212	5421	4706	3592	2777	1925	1085	599	352	188	131	64	39	582.02
ZYX 2NN	102	127	216	307	525	791	1153	1667	2238	3055	4127	5981	4717	3160	1715	482	104	14	3	1	0	0	3541.18
ZYX 3NN	113	136	248	339	580	845	1217	1664	2068	2704	3740	6344	4915	3185	1479	320	51	7	2	0	0	0	5162.35
ZYX 2N2N	39	58	152	233	399	655	1102	1899	2700	3537	4654	5455	4269	3204	1984	798	276	94	13	6	2	1	712.77
ZXA 2NN	18	31	86	110	218	474	975	1314	1657	4460	6663	3505	3021	2787	2113	937	531	386	247	168	69	45	7091.82
ZXA 3NN	45	66	109	155	350	677	1037	1358	1801	4040	5757	4277	3856	3012	1702	879	531	302	150	93	37	29	3923.93
ZXA 2N2N	0	1	11	21	79	229	527	1743	2558	4868	6587	3472	3076	2755	2131	1107	620	434	294	204	96	76	4735.63
ZAX 2NN	127	149	227	340	656	1024	1412	1741	1861	1914	2461	7622	4778	2595	1097	235	62	15	4	1	0	0	11295.80
ZAX 3NN	134	150	235	357	678	1056	1474	1755	1877	1929	2342	7703	5407	2722	822	111	15	4	1	0	0	0	12490.21
ZAX 2N2N	59	83	220	301	451	631	1109	2124	2733	3067	3421	6631	4859	2627	1808	488	210	87	27	14	2	2	4459.02
ZXB 2NN	0	0	0	5	51	181	603	1883	3319	4609	5662	4487	3548	2677	1873	1122	656	445	276	193	89	64	1632.03
ZXB 3NN	0	0	6	33	114	296	802	1997	3121	4246	5319	4732	3726	2748	1895	1041	632	392	215	145	62	36	571.31
ZXB 2N2N	0	0	0	2	12	66	417	1735	3448	4954	6017	4041	3331	2688	1993	1155	676	442	302	216	102	71	2811.77
ZBX 2NN	125	148	254	359	636	975	1385	1753	1907	2146	3006	7059	5379	3074	990	53	2	1	0	0	0	0	9194.15
ZBX 3NN	134	155	254	360	639	1036	1459	1771	1889	2036	2648	7386	5726	3039	749	22	3	0	0	0	0	0	11836.09
ZBX 2N2N	74	111	225	312	488	736	1165	1953	2607	3241	4030	6037	4994	3505	1778	389	46	6	1	1	0	0	3077.23
ZYA 2NN	53	70	146	208	363	571	976	1794	2714	3754	4789	5362	4323	3095	1926	850	343	151	31	9	3	2	749.70
ZYA 3NN	70	85	181	255	393	651	1015	1728	2562	3504	4515	5552	4337	3087	1893	743	299	121	25	8	2	0	1366.43
ZYA 2N2N	19	28	80	138	294	546	1007	1919	2874	3878	4888	5194	4145	3062	1974	942	416	196	54	18	4	2	106.84
ZAY 2NN	3	8	44	83	236	468	969	2038	3033	4055	5114	4948	3914	2965	2021	985	523	284	129	73	22	12	40.16
ZAY 3NN	3	9	51	87	242	475	1012	2018	3037	4033	5055	4981	3919	2968	1987	1005	506	264	117	66	19	11	25.09
ZAY 2N2N	0	5	27	71	217	487	983	1981	2992	4059	5110	4967	3921	2853	1933	1006	512	265	134	74	29	23	106.88
ZX NN	56	67	124	150	323	689	1197	1589	1653	2065	5428	5300	2104	1659	1565	1206	702	346	159	124	68	52	11600.22
ZY NN	7	14	49	102	267	501	999	2093	3074	4115	5148	5017	4024	3038	2061	1060	542	265	96	50	6	3	35.29

N= 4

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZYY 2NN	1	10	41	104	250	503	1012	2014	3052	4017	5044	5017	4028	3046	2021	1035	514	265	121	65	13	5	24.91
ZYY 3NN	3	9	53	100	241	479	1013	2031	3001	3976	4966	5082	4034	3020	1987	1021	529	242	118	50	15	4	35.65
ZYY 2N2N	8	20	63	119	264	524	1025	2011	2993	4022	5086	5016	3949	2948	2018	1044	548	286	126	70	17	6	34.10
ZXX 2NN	46	68	157	232	354	564	1103	1946	2481	2988	4581	5563	3999	3040	1625	748	302	155	66	37	10	6	1754.18
ZXX 3NN	59	89	175	245	399	620	1108	1953	2554	3027	4106	5972	4584	3046	1641	563	199	84	29	16	2	1	2037.51
ZXX 2N2N	17	22	67	119	267	514	940	1944	2941	3639	5087	5063	3661	2938	1921	933	516	272	125	80	30	18	515.55
ZXY 2NN	1	3	26	59	192	410	935	1974	3142	4206	5211	4876	3928	2950	1944	1014	549	310	146	98	45	26	211.09
ZXY 3NN	3	5	41	81	205	441	953	1987	3047	4058	5082	4972	3947	2924	1940	1003	497	280	138	83	25	15	58.68
ZXY 2N2N	0	1	16	35	150	374	880	2029	3136	4231	5351	4731	3664	2759	1903	1078	600	356	195	123	45	34	387.06
ZYX 2NN	57	73	178	249	432	654	1104	1877	2662	[illegible]64	4368	5706	4589	3287	1900	657	207	54	6	1	0	0	1356.43
ZYX 3NN	64	83	183	260	436	675	1126	1911	2612	3283	4142	5890	4809	3422	1852	495	108	17	1	0	0	0	2069.02
ZYX 2N2N	32	45	120	197	360	606	1064	1960	2845	3734	4799	5291	4209	3152	2087	911	388	163	37	20	1	1	333.51
ZXA 2NN	8	12	49	98	210	377	938	1707	2495	4625	5605	4444	3833	2798	1726	1035	568	296	168	109	49	38	1940.89
ZXA 3NN	17	29	95	145	260	485	1010	1700	2624	4056	4935	5086	3893	2705	1811	839	451	238	121	79	31	18	541.18
ZXA 2N2N	0	2	12	26	85	247	721	1954	3149	4739	5580	4453	3712	2732	1824	1114	651	373	224	150	62	49	1257.83
ZAX 2NN	71	103	210	313	450	623	1101	2046	2760	3141	3438	6579	5348	3346	1423	293	77	20	4	0	0	0	4630.10
ZAX 3NN	82	106	217	316	455	625	1106	2106	2794	3195	3438	6568	5732	3642	1329	161	35	7	1	0	0	0	6174.25
ZAX 2N2N	38	50	131	204	395	664	1096	1847	2764	3732	4446	5601	4758	3113	1946	695	245	102	29	7	1	0	1150.41
ZXB 2NN	0	0	3	12	72	263	799	2049	3275	4454	5444	4644	3663	2769	1931	1046	631	379	203	140	64	50	827.52
ZXB 3NN	0	1	11	38	135	336	865	2018	3130	4233	5260	4782	3757	2809	1895	1013	558	332	177	117	56	35	392.68
ZXB 2N2N	0	0	0	2	30	157	681	2060	3448	4624	5576	4469	3569	2702	1892	1090	662	433	245	169	72	58	1389.61
ZBX 2NN	75	104	222	310	454	650	1135	2029	2742	3187	3724	6309	5375	3725	1580	223	12	0	0	0	0	0	4431.48
ZBX 3NN	81	104	225	302	475	656	1103	2088	2771	3216	3564	6448	5676	3904	1454	102	3	0	0	0	0	0	5961.28
ZBX 2N2N	46	64	162	245	420	674	1120	1879	2726	3601	4525	5520	4612	3443	2121	656	168	34	1	1	0	0	1197.74
ZYA 2NN	22	32	100	173	326	582	1061	1969	2827	3847	4873	5239	4170	3046	2026	918	426	192	54	24	2	1	176.21
ZYA 3NN	38	68	127	193	363	605	1035	1899	2758	3652	4680	5368	4219	3147	1980	872	330	128	30	13	0	0	469.43
ZYA 2N2N	16	26	92	151	311	550	1040	1995	2922	3947	4981	5090	4065	3019	2100	1010	498	241	97	31	9	1	92.88
ZAY 2NN	4	6	43	83	229	480	980	1973	3029	4023	5058	4979	3948	3017	2020	1013	504	258	107	67	19	9	24.84
ZAY 3NN	4	8	48	93	240	500	964	1988	2994	4020	5018	5015	3974	2984	2000	1000	502	251	116	54	15	8	13.56
ZAY 2N2N	2	7	35	84	229	470	977	1994	3001	4089	517	4883	3854	2824	1938	1050	571	299	138	80	24	14	76.34
ZX NN	21	30	93	157	298	434	957	1991	2423	2833	5195	5169	2804	2396	1924	897	418	281	161	90	38	24	4377.65
ZY NN	1	5	41	98	229	496	1047	2035	3038	3991	5082	5034	3927	3028	2090	1025	508	255	98	50	12	7	39.59

N= 8

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES												
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZYY 2NN	3	12	48	110	265	501	977	1949	3016	3999	5001	5044	3993	3026	1998	1025	505	231	93	49	9	5	23.90
ZYY 3NN	7	12	56	104	267	502	1008	1986	2998	4047	5069	4958	4010	3012	2030	996	521	241	103	54	9	5	18.91
ZYY 2N2N	9	18	58	108	245	483	964	1927	2921	3905	5013	5052	3954	2998	2059	1009	485	255	93	39	7	1	37.67
ZXX 2NN	40	51	136	197	364	[illegible]22	1046	1894	2817	3620	4614	5422	4333	3094	1885	796	339	142	55	22	7	3	564.96
ZXX 3NN	49	63	159	220	398	651	1089	1925	2787	3633	4492	5540	4475	3275	1978	707	223	74	24	10	0	0	962.13
ZXX 2N2N	13	23	62	105	254	512	974	1909	2906	3843	4946	5119	3976	2973	1932	943	483	261	120	68	12	7	63.31
ZXY 2NN	0	3	31	72	208	454	972	1997	3020	4058	5076	4981	3933	2945	1995	1042	539	303	126	74	22	11	44.72
ZXY 3NN	4	10	40	83	233	468	1014	2012	3014	4045	5064	4971	3957	2983	2006	1030	527	293	123	61	14	9	22.79
ZXY 2N2N	1	4	15	55	153	391	927	1936	2985	4018	5151	4912	3839	2904	1960	1031	556	324	164	102	28	15	144.01
ZYX 2NN	52	69	150	230	383	617	1062	1887	2732	3662	4598	5445	4417	3257	1990	742	230	72	10	2	0	0	930.64
ZYX 3NN	55	75	168	247	405	657	1101	1883	2750	3610	4512	5517	4492	3341	2014	659	163	37	3	0	0	0	1237.52
ZYX 2N2N	18	36	109	167	325	572	1015	1892	2842	3803	4751	5308	4191	3088	1936	886	377	148	51	22	2	1	211.76
ZXA 2NN	9	14	48	89	218	439	868	1890	2994	4150	5108	4956	3892	2864	1896	1004	538	307	155	97	28	20	150.60
ZXA 3NN	20	28	82	134	282	502	974	1893	2920	3938	4982	5039	3966	2945	1945	950	451	246	103	57	14	11	87.06
ZXA 2N2N	0	0	13	31	106	320	827	1999	3166	4202	5321	4719	3733	2843	1934	1058	622	372	197	127	51	29	418.76
ZAX 2NN	65	84	175	254	438	689	1138	1853	2677	3588	4476	5552	4764	3677	1970	475	96	23	4	1	0	0	2153.21
ZAX 3NN	65	90	183	262	428	703	1189	1881	2633	3583	4532	5407	4740	3877	2098	372	40	7	0	0	0	0	2756.89
ZAX 2N2N	29	48	121	184	383	635	1065	1899	2745	3619	4621	5410	4353	3282	1998	784	310	112	25	10	2	2	506.90
ZXB 2NN	0	0	5	32	123	351	853	2035	3147	4229	5245	4834	3827	2839	1943	1063	587	370	182	121	36	26	317.10
ZXB 3NN	0	0	23	51	172	393	929	2031	3065	4153	5180	4842	3866	2914	197[illegible]	1036	563	331	150	82	29	21	141.87
ZXB 2N2N	0	0	2	9	71	265	820	1965	3181	4297	5344	4694	3720	2809	1954	1070	630	390	210	132	50	34	545.43
ZBX 2NN	64	88	187	260	434	681	1138	1859	2657	3543	4434	5585	467[illegible]	3523	2104	415	53	2	0	0	0	0	2219.71
ZBX 3NN	61	88	187	272	439	710	1176	1867	2650	3546	4526	5492	4688	3791	2207	330	22	0	0	0	0	0	2690.24
ZBX 2N2N	36	59	140	213	391	644	1046	1880	2758	3616	4586	5454	4397	3255	2044	784	243	75	3	1	0	0	697.90
ZYA 2NN	21	31	113	173	309	526	990	1927	2913	3878	4868	5207	4156	3062	1995	910	388	188	69	20	2	2	175.23
ZYA 3NN	27	46	125	187	346	562	1020	1946	2845	3832	4772	5254	4184	3071	1974	868	357	146	38	18	2	0	300.72
ZYA 2N2N	12	21	68	122	272	525	993	1956	2994	3925	4916	5123	4119	3044	1986	969	478	222	80	37	2	2	43.67
ZAY 2NN	5	10	44	94	240	490	1011	1971	2979	3976	5063	4964	3987	2984	2008	1055	502	266	123	66	11	4	27.42
ZAY 3NN	6	13	43	96	251	504	1019	1990	3025	4045	5005	5005	3998	2978	2032	1023	522	270	114	55	9	8	20.00
ZAY 2N2N	4	6	38	83	225	458	963	1946	2943	3933	4989	5045	3999	2984	1999	1022	498	247	112	63	17	9	18.05
ZX NN	24	29	87	134	273	524	933	1885	2896	3587	5005	5134	3686	2942	1891	927	492	253	117	62	18	10	547.85
ZY NN	4	8	46	98	267	517	1035	2020	2985	3974	4995	5098	4019	3013	2034	999	488	235	92	45	11	[illegible]	11.30

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES										RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZYY 2NN	5	9	47	86	255	515	1005	2004	3035	4057	5112	4928	3928	2922	1950	969	443	215	80	37	10	5	19.97
ZYY 3NN	7	11	50	102	251	494	1001	2012	3043	4089	5050	4978	3905	2909	1954	974	462	215	73	41	11	5	23.27
ZYY 2N2N	4	10	43	90	235	492	1014	2015	2973	4010	5065	4976	3917	2894	1925	940	445	214	88	49	9	4	18.61
ZXX 2NN	30	42	116	182	365	629	1056	1946	2840	3808	4805	5229	4196	3090	1936	876	358	175	64	21	1	0	299.01
ZXX 3NN	34	52	133	198	374	622	1071	1933	2791	3728	4712	5309	4312	3145	1986	842	328	122	35	10	1	1	455.54
ZXX 2N2N	4	8	61	111	265	[illegible]	997	2041	3010	4042	5121	4936	3874	2907	1979	983	459	253	107	50	8	2	31.42
ZXY 2NN	4	9	33	71	233	507	992	2026	3131	4172	5148	4877	3883	2940	1943	1001	491	267	118	61	12	7	42.07
ZXY 3NN	7	9	29	87	242	488	987	2021	3055	4118	5139	4883	3900	2952	2000	1000	504	256	101	51	12	6	26.93
ZXY 2N2N	1	4	24	60	195	452	969	2064	3145	4226	5156	4895	3881	2862	1940	1022	543	299	141	78	24	15	97.70
ZYX 2NN	30	48	118	182	355	638	1095	1959	2848	3782	4769	5268	4244	3171	2018	831	302	90	16	9	0	0	413.67
ZYX 3NN	39	59	136	196	373	643	1105	1940	2838	3736	4682	5345	4285	3193	2035	817	269	83	11	4	0	0	598.41
ZYX 2N2N	9	17	68	125	281	553	1039	2012	2958	3935	4972	5077	4033	2991	1957	896	412	180	35	17	2	1	67.40
ZXA 2NN	3	8	36	104	226	487	1005	2002	3075	4129	5143	4891	3872	2911	1895	955	506	295	134	85	21	9	70.40
ZXA 3NN	16	25	69	122	285	533	1008	1980	2966	3935	4997	5024	3950	2945	1910	902	461	248	116	54	16	8	60.36
ZXA 2N2N	0	0	10	40	149	372	941	2077	3262	4339	5303	4720	3765	2801	1911	1001	568	336	163	94	24	18	227.26
ZAX 2NN	49	64	154	224	404	669	1127	1931	2781	3620	4543	5476	4378	3284	2077	798	209	55	2	0	0	0	931.32
ZAX 3NN	47	73	148	225	395	663	1111	1928	2750	3617	4528	5487	4441	3316	2130	773	153	26	2	0	0	0	1058.14
ZAX 2N2N	13	28	116	175	357	581	1051	1932	2905	3895	4872	5159	4146	3114	2021	912	372	143	37	19	0	0	200.76
ZXB 2NN	1	2	17	42	157	422	981	2109	3231	4260	5297	4759	3784	2850	1909	1013	532	311	154	94	23	13	152.44
ZXB 3NN	2	6	23	55	185	439	973	2023	3148	4200	5205	4814	3816	2883	1931	1004	525	291	130	82	21	11	85.87
ZXB 2N2N	0	0	6	23	133	358	906	2130	3275	4310	5321	4709	3731	2786	1907	1021	570	337	188	107	32	18	288.74
ZBX 2NN	48	73	145	227	388	662	1127	1930	2745	3628	4625	5396	4396	3326	2093	767	164	23	0	0	0	0	1030.19
ZBX 3NN	47	74	155	222	389	677	1126	1921	2761	3638	4577	5433	4426	3343	2133	788	123	8	0	0	0	0	1129.21
ZBX 2N2N	22	33	103	177	349	592	1063	1959	2873	3863	4844	5183	4197	3154	2014	884	339	112	18	4	0	0	257.99
ZYA 2NN	8	12	66	122	285	594	1081	1986	2975	3956	4954	5085	4044	2975	1943	925	403	178	54	16	5	1	74.83
ZYA 3NN	17	24	93	160	327	584	1096	2012	2920	3865	4885	5132	4076	3041	1967	893	384	172	50	23	2	2	129.50
ZYA 2N2N	3	9	49	102	252	490	1029	2018	3066	4031	5092	4934	3974	2918	1897	924	469	217	67	28	5	2	34.18
ZAY 2NN	7	12	45	91	240	492	977	1972	3061	4099	5103	4916	3930	2935	1980	1010	486	237	93	50	13	6	19.34
ZAY 3NN	6	12	41	94	244	485	976	2016	3044	4074	5086	4936	3944	2940	1994	1031	492	246	92	44	9	4	16.16
ZAY 2N2N	4	7	44	92	251	496	990	2041	3034	4068	5079	4950	3976	2949	1941	1004	516	277	121	60	20	:0	25.67
ZX NN	9	23	85	142	280	514	991	1954	2948	3937	5056	5007	3904	2920	1912	965	489	288	128	70	19	8	77.67
ZY NN	5	10	43	97	250	500	1031	2061	2980	3982	5076	5000	3910	2920	1942	980	470	225	94	43	9	5	22.69

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZYY 2NN	3	8	43	81	223	472	978	1981	2937	3949	4982	5041	4016	2993	1995	1011	512	265	105	56	15	10	14.08
ZYY 3NN	6	10	40	90	237	477	987	1947	2958	3972	4977	5044	4062	2992	2027	1024	526	288	118	58	12	7	19.40
ZYY 2N2N	5	5	42	93	230	454	979	1988	2977	3854	4877	5064	3939	2993	2020	988	498	259	106	55	15	7	30.42
ZXX 2NN	8	13	66	115	285	550	1022	1904	2828	3817	4834	5191	4175	3065	1958	894	405	198	71	34	8	5	78.92
ZXX 3NN	13	20	74	144	318	578	1047	1938	2862	3764	4720	5300	4275	3134	2004	878	383	176	43	17	4	1	162.71
ZXX 2N2N	5	9	44	91	223	452	947	1951	2910	3957	4973	5056	4013	2987	1964	939	464	239	105	58	20	12	28.87
ZXY 2NN	3	3	24	68	215	463	957	1957	2998	4031	5075	4949	3923	2996	2012	1024	537	293	132	68	14	9	38.72
ZXY 3NN	2	5	32	78	215	488	1011	1974	2995	4000	5002	5012	3971	3012	2043	1022	541	293	134	73	16	6	33.67
ZXY 2N2N	0	3	30	62	173	420	926	1997	3029	4014	5070	4964	3901	2886	1967	1008	536	277	131	79	24	14	74.05
ZYX 2NN	13	22	72	139	301	552	1031	1923	2819	3754	4756	5271	4210	3066	1977	898	388	169	52	21	2	0	126.47
ZYX 3NN	14	27	99	154	316	567	1032	1945	2807	3745	4753	5265	4279	3158	2001	897	371	153	36	15	0	0	194.76
ZYX 2N2N	8	15	61	108	265	496	1000	1944	2877	3838	4869	5160	4092	3007	1985	943	471	201	75	31	5	0	44.86
ZXA 2NN	1	3	27	61	211	453	951	1973	2973	4021	5037	4991	3962	2865	1895	942	493	257	122	72	20	12	55.00
ZXA 3NN	2	7	39	93	258	506	991	1992	2978	3920	4997	5016	3992	2961	1979	929	457	229	95	67	13	9	27.90
ZXA 2N2N	0	0	17	54	168	423	911	1974	3047	4129	5167	4855	3871	2856	1911	937	522	290	136	79	32	21	133.50
ZAX 2NN	14	26	100	162	328	595	1039	1896	2769	3654	4628	5384	4288	3198	2059	888	332	118	22	6	1	0	275.34
ZAX 3NN	22	31	94	179	349	599	1056	1925	2747	3654	4630	5380	4338	3294	2082	871	323	97	15	3	0	0	367.04
ZAX 2N2N	13	24	61	120	267	495	989	1931	2836	3802	4774	5249	4183	3151	2033	908	401	173	64	27	2	0	101.61
ZXB 2NN	0	2	20	49	178	418	907	2018	3083	4158	5145	4906	3876	2912	1916	1017	538	285	143	91	26	17	115.54
ZXB 3NN	1	6	25	58	197	445	993	2017	3076	4108	5091	4930	3919	2935	1959	1024	539	292	131	77	20	13	58.77
ZXB 2N2N	0	0	15	47	152	370	924	1996	3102	4139	5173	4848	3814	2873	1891	993	546	300	165	97	33	21	177.70
ZBX 2NN	17	28	112	166	342	569	1067	1915	2752	3660	4627	5388	4322	3205	2023	886	341	114	15	4	0	0	316.46
ZBX 3NN	23	35	112	173	352	593	1076	1898	2749	3635	4620	5389	4342	3272	2106	868	331	93	11	1	0	0	390.09
ZBX 2N2N	11	21	69	130	274	540	1008	1884	2811	3792	4774	5245	4186	3142	2043	933	399	177	46	13	0	0	109.50
ZYA 2NN	7	12	63	101	249	496	996	1945	2913	3906	4935	5092	4026	2974	1951	947	447	211	78	34	8	4	25.42
ZYA 3NN	12	20	71	121	274	502	991	1953	2943	3896	4906	5112	4095	3015	1935	927	451	197	72	28	2	1	50.95
ZYA 2N2N	7	11	47	98	242	479	953	1999	2961	3955	4966	5053	4010	2983	1950	951	491	243	73	39	9	3	24.73
ZAY 2NN	2	3	33	77	235	496	948	1962	2966	3953	4993	5018	4039	3029	2051	1047	523	288	125	65	15	9	27.69
ZAY 3NN	3	5	39	89	242	493	975	1974	2992	3990	4979	5032	4053	3054	2086	1038	532	275	126	76	11	8	27.34
ZAY 2N2N	3	7	29	76	198	468	950	1977	2963	3974	4978	5042	4034	2980	2027	1015	534	278	123	69	15	11	36.76
ZX NN	6	9	47	97	233	483	983	1934	2913	3961	5021	5017	3998	2939	1943	982	471	242	97	50	10	8	20.20
ZY NN	6	11	42	86	242	469	943	1955	2968	3953	5028	5034	4007	2983	2030	1049	520	257	111	54	8	5	17.43

N= 64

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES										RIGHT-TAIL CUMULATIVE PROBABILITIES												
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZYY 2NN	11	13	56	111	289	530	1035	1996	2969	3947	4975	5039	4023	2992	1941	959	457	220	91	43	4	1	30.20
ZYY 3NN	8	15	65	118	261	533	1037	2032	2979	3951	4989	5024	4005	2997	1956	975	477	238	98	35	9	2	26.49
ZYY 2N2N	4	9	47	107	255	492	995	2029	3010	3985	5000	5020	4017	2975	1969	967	469	235	92	44	7	5	10.40
ZXX 2NN	9	13	76	148	322	587	1064	2036	2984	3946	4958	5061	4037	3000	1931	926	417	197	70	28	4	2	61.98
ZXX 3NN	10	26	86	162	354	609	1109	2012	3013	3941	4917	5093	4124	3114	1994	944	404	165	55	24	4	0	135.08
ZXX 2N2N	6	10	39	85	241	508	1027	2057	3085	4088	5093	4929	3915	2925	1936	938	448	210	91	55	9	3	20.83
ZXY 2NN	3	3	43	92	241	519	1030	2037	3049	4055	5070	4944	3966	2946	1955	970	479	264	101	50	8	2	19.35
ZXY 3NN	6	11	54	99	264	524	1029	2045	3065	4045	5021	5001	3938	2956	1975	1008	485	236	109	49	6	3	18.05
ZXY 2N2N	2	3	32	75	207	437	974	2046	3101	4124	5112	4905	3890	2855	1926	1012	500	247	101	62	19	10	49.64
ZYX 2NN	15	24	86	160	342	596	1073	2019	2942	3874	4852	5164	4108	3028	2027	914	411	173	48	22	2	1	123.29
ZYX 3NN	15	29	104	174	344	651	1106	1995	2928	3839	4838	5174	4117	3078	2010	915	409	164	48	19	0	0	174.81
ZYX 2N2N	3	9	57	118	290	557	1039	1999	3019	4020	4960	5064	3993	2982	2013	964	433	207	74	29	4	3	40.48
ZXA 2NN	6	7	43	95	243	501	1032	2055	3091	4128	5167	4858	3844	2908	1907	969	501	267	112	69	15	11	34.44
ZXA 3NN	8	13	53	117	294	545	1070	2073	3105	4082	5060	4946	3955	2995	1998	981	486	227	100	57	15	10	26.04
ZXA 2N2N	1	1	30	58	177	438	999	2114	3188	4210	5175	4839	3791	2807	1883	999	512	276	132	77	25	15	107.82
ZAX 2NN	19	33	104	167	374	621	1084	2024	2914	3878	4813	5200	4174	3102	2018	900	375	143	32	10	0	0	216.76
ZAX 3NN	23	32	112	186	362	623	1101	2005	2930	3895	4801	5207	4216	3173	2026	913	356	122	31	12	0	0	267.09
ZAX 2N2N	9	16	53	120	277	569	1055	2010	2961	3973	4954	5056	4082	3032	1965	917	399	178	51	25	2	1	65.86
ZXB 2NN	1	3	30	71	217	474	997	2070	3150	4160	5184	4837	3860	2858	1924	962	516	265	126	74	16	9	51.07
ZXB 3NN	3	5	43	94	240	510	1002	2045	3138	4138	5078	4931	3931	2950	1977	1007	510	258	123	66	20	10	34.64
ZXB 2N2N	0	1	25	59	178	417	958	2086	3165	4219	5210	4806	3808	2830	1906	1022	544	277	134	85	27	16	117.59
ZBX 2NN	23	38	104	180	365	621	1094	2001	2897	3787	4767	5240	4183	3089	2006	891	356	134	33	12	1	0	256.56
ZBX 3NN	23	36	110	181	371	637	1102	1990	2912	3824	4788	5219	4219	3135	2034	921	376	125	33	8	0	0	262.01
ZBX 2N2N	7	14	75	123	315	567	1071	1980	2953	3956	4863	5154	4112	3028	1995	920	404	189	45	20	3	0	95.90
ZYA 2NN	9	18	77	138	294	559	1060	2014	3034	3982	4991	5032	3988	3006	2016	942	444	203	74	42	7	4	42.55
ZYA 3NN	8	17	79	131	319	596	1108	2000	2965	3966	4947	5067	4030	3000	1989	955	445	215	68	34	5	1	60.06
ZYA 2N2N	7	11	59	127	273	535	1011	2020	3021	4072	5035	4982	3997	2996	1999	975	454	214	87	43	9	7	25.83
ZAY 2NN	5	9	60	106	251	527	1034	2036	3001	3997	5015	4998	3967	2989	1969	967	473	246	82	36	7	3	17.22
ZAY 3NN	5	13	60	114	257	518	1023	2029	2968	3977	5008	4996	3984	2990	1981	993	499	234	88	45	6	3	12.95
ZAY 2N2N	2	6	44	89	223	468	974	2008	3029	4035	5059	4950	3967	2918	1946	982	476	238	93	48	8	7	16.75
ZX NN	7	10	59	106	270	526	1009	2036	3051	4025	5071	4951	3962	2935	1977	961	494	249	112	56	5	3	21.67
ZY NN	4	12	58	104	237	493	994	2016	2945	3980	4986	5046	3982	2971	1960	994	481	239	86	42	8	4	18.52

N= 128

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZYY 2NN	3	10	55	109	240	482	953	1912	2956	3937	4954	5055	4050	3080	2062	1041	494	247	87	42	7	3	19.93
ZYY 3NN	1	7	52	104	239	474	964	1955	2976	3958	4896	5110	4064	3080	2106	1017	503	226	87	45	5	2	30.47
ZYY 2N2N	4	7	46	99	229	496	988	1921	2948	3947	4971	5049	4009	3031	2005	993	488	233	87	43	8	3	15.55
ZXX 2NN	10	15	74	136	270	488	948	1877	2829	3830	4807	5202	4153	3127	2065	964	470	225	72	30	4	2	60.72
ZXX 3NN	11	22	81	134	267	499	953	1891	2810	3778	4778	5229	4191	3107	2092	961	439	195	64	25	1	0	88.67
ZXX 2N2N	4	9	55	99	263	506	977	1927	2921	3945	4924	5086	4039	3026	2046	1044	513	247	79	33	9	4	23.16
ZXY 2NN	5	10	51	105	247	456	952	1927	2958	3989	4955	5058	4100	3100	2058	1020	494	249	100	47	9	6	20.01
ZXY 3NN	4	10	53	111	221	481	961	1912	2954	3973	4944	5069	4131	3101	2082	1001	490	249	92	41	9	6	35.58
ZXY 2N2N	3	7	39	89	237	508	981	1936	2933	3955	4987	5026	3998	2971	2032	1048	532	261	108	50	10	2	21.40
ZYX 2NN	11	14	62	132	291	502	960	1934	2916	3850	4833	5181	4122	3141	2105	1036	466	208	63	21	4	1	71.53
ZYX 3NN	10	24	72	143	277	487	958	1900	2894	3826	4808	5206	4171	3116	2070	1019	461	203	61	19	2	1	86.90
ZYX 2N2N	7	8	50	128	286	515	965	1955	2905	3924	4920	5095	4059	3081	2072	997	495	230	85	32	5	4	45.95
ZXA 2NN	6	10	44	91	231	447	934	1930	2918	3911	4890	5125	4092	3029	2088	1003	521	268	100	50	9	4	26.52
ZXA 3NN	7	11	60	113	236	459	929	1879	2865	3851	4859	5149	4099	3091	2045	1028	489	232	93	52	7	3	27.56
ZXA 2N2N	2	6	38	79	210	447	968	1938	2989	3946	4995	5018	3995	3030	2049	1055	546	273	117	60	11	6	23.96
ZAX 2NN	11	31	88	151	295	507	997	1871	2822	3788	4777	5231	4181	3159	2028	967	428	168	50	11	0	0	155.87
ZAX 3NN	14	31	85	145	290	518	1005	1880	2827	3778	4762	5246	4199	3144	2099	952	425	176	46	11	0	0	143.74
ZAX 2N2N	10	16	70	138	286	536	983	1980	2888	3635	4868	5142	4061	3039	2032	1026	477	207	59	21	4	0	68.37
ZXB 2NN	3	6	38	87	224	476	944	1945	2968	3973	4987	5027	3998	3065	2069	1074	537	275	130	56	11	5	27.85
ZXB 3NN	2	10	49	88	222	457	959	1904	2883	3931	4957	5047	4063	3045	2105	1022	553	271	102	52	10	4	35.07
ZXB 2N2N	3	5	31	69	220	455	946	1987	2972	4011	5045	4966	3975	3024	2045	1066	563	294	120	65	13	6	29.58
ZBX 2NN	14	25	81	149	297	533	1006	1897	2858	3813	4784	5223	4178	3142	2087	981	440	177	50	16	0	0	109.40
ZBX 3NN	14	28	88	147	302	536	997	1905	2842	3795	4800	5205	4182	3146	2098	989	443	187	48	7	0	0	122.83
ZBX 2N2N	8	19	78	136	289	538	1000	1931	2885	3891	4893	5122	4127	3078	2074	1001	462	200	60	25	2	0	59.91
ZYA 2NN	6	11	57	113	240	476	968	1949	2904	3891	4915	5103	4096	3114	2101	1012	505	231	91	44	5	4	24.13
ZYA 3NN	6	16	69	117	255	455	928	1915	2878	3872	4887	5116	4089	3131	2117	1039	492	244	85	34	4	3	43.87
ZYA 2N2N	6	11	52	105	262	484	949	1968	2951	3890	4912	5099	4081	3079	2028	1005	517	240	96	48	7	6	21.20
ZAY 2NN	2	14	59	107	237	479	966	1965	2973	3963	4904	5106	4089	3078	2083	1013	465	243	90	42	11	7	35.93
ZAY 3NN	4	9	60	110	229	486	967	1962	2989	3982	4902	5107	4116	3114	2059	1012	465	254	82	44	9	7	43.75
ZAY 2N2N	4	12	55	103	269	496	991	1956	2930	3960	4963	5047	3986	2999	2024	1001	499	250	87	44	7	3	16.79
ZX NN	4	10	52	102	241	477	980	1951	2952	3937	4933	5084	4089	3099	2041	994	502	250	94	38	11	8	17.12
ZY NN	7	10	50	91	241	488	970	1966	2946	3942	4917	5105	4047	3048	2060	1022	513	244	78	44	9	4	18.69

N= 256

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZYY 2NN	3	9	46	102	234	480	1009	2005	3005	4010	5048	4960	3995	2953	1954	986	490	254	103	42	4	3	18.70
ZYY 3NN	4	7	55	102	255	481	997	2013	2990	4036	5063	4943	3940	2959	1986	959	489	256	101	43	5	3	18.05
ZYY 2N2N	7	14	51	97	258	487	979	1984	3011	4047	5051	4961	4036	3014	2031	1014	506	239	96	46	7	4	17.16
ZXX 2NN	10	16	60	116	267	535	1046	1982	2945	3921	4916	5088	4048	3072	2050	958	449	214	77	40	10	3	34.59
ZXX 3NN	14	22	73	130	285	531	1023	1970	2945	3908	4897	5111	4073	3063	1991	964	459	208	72	42	6	3	45.92
ZXX 2N2N	4	9	44	89	265	534	1018	1991	2928	4015	5030	4983	4013	3014	2039	992	512	248	84	37	2	1	34.39
ZXY 2NN	3	10	50	97	237	469	977	2004	3061	4067	5046	4959	3956	2974	2012	1036	511	258	106	57	10	5	13.34
ZXY 3NN	5	9	53	95	237	464	964	1993	3086	3990	5057	4948	3940	2962	2018	1012	522	259	102	53	12	6	32.62
ZXY 2N2N	2	7	40	91	232	497	977	2024	2988	4013	4999	5013	3990	2996	2031	1054	550	263	111	56	12	10	25.13
ZYX 2NN	8	14	67	131	282	564	1032	1988	2935	3919	4932	5078	4090	3037	1989	950	439	200	62	32	4	1	45.57
ZYX 3NN	11	19	78	134	291	556	1022	2017	2942	3922	4950	5052	4080	3023	1962	937	453	212	63	25	3	0	55.12
ZYX 2N2N	6	12	59	117	270	524	1006	1991	2998	3983	5008	4999	3995	2982	1991	980	458	219	84	42	5	2	12.07
ZXA 2NN	6	8	35	92	253	483	996	1983	3005	4035	5007	5006	3989	3004	2012	1025	516	241	97	48	7	5	18.01
ZXA 3NN	7	14	61	115	245	485	991	1980	2940	3946	4975	5027	4021	3010	2002	976	483	251	104	56	8	6	14.84
ZXA 2N2N	3	5	28	72	221	478	1018	1980	3010	4031	5062	4944	3959	2974	2035	1042	544	300	106	44	8	5	38.97
ZAX 2NN	15	25	79	134	287	564	1061	1986	2920	3919	4897	5107	4114	3077	2053	946	427	188	65	29	4	2	76.64
ZAX 3NN	14	27	80	139	302	538	1060	1970	2969	3909	4885	5117	4137	3085	1975	954	443	177	54	32	4	1	95.14
ZAX 2N2N	10	15	57	114	309	562	1033	2028	2968	3935	4911	5094	4094	3061	2004	960	448	223	73	23	2	1	51.21
ZXB 2NN	3	4	30	79	224	464	971	2014	3055	4070	5070	4936	3977	2994	2005	1049	537	267	108	61	13	7	22.84
ZXB 3NN	4	10	43	92	225	466	973	1992	3055	4063	5018	4987	3996	2980	2019	1007	508	277	118	63	16	10	21.61
ZXB 2N2N	0	2	22	72	209	462	969	2011	3029	4015	5021	4984	3989	2962	2028	1076	566	289	117	62	14	10	41.44
ZBX 2NN	13	20	75	134	310	585	1070	1966	2959	3920	4873	5134	4142	3047	2007	918	422	183	52	23	1	1	89.96
ZBX 3NN	13	24	82	141	302	561	1065	1995	2947	3915	4858	5144	4092	3082	1959	937	442	180	55	24	3	0	87.63
ZBX 2N2N	9	23	65	124	280	549	1042	1998	2970	3970	4917	5090	4059	3023	1988	974	426	208	75	35	2	0	52.94
ZYA 2NN	3	12	54	106	264	509	1008	1993	2960	3971	4963	5054	4023	2969	1952	969	466	217	80	37	7	4	16.04
ZYA 3NN	5	11	64	116	297	518	1012	1979	2999	4005	4994	5011	3994	3018	1939	932	465	235	84	43	3	2	33.54
ZYA 2N2N	7	9	47	87	254	512	988	1993	3023	4008	5023	4983	3956	2958	1954	1009	497	240	97	41	9	5	16.28
ZAY 2NN	6	10	55	98	249	484	994	1994	3036	4019	5071	4932	3964	2963	2001	1039	510	251	103	53	10	5	13.98
ZAY 3NN	6	12	54	94	267	498	993	1990	3031	4033	5059	4943	3940	2948	2020	1022	507	252	102	43	10	8	22.09
ZAY 2N2N	8	13	56	98	265	476	986	1994	3042	4024	5007	5003	4036	3066	2043	1039	508	255	102	61	17	9	26.62
ZX NN	5	8	60	109	254	513	1013	1999	2938	3944	4943	5070	4009	2946	1942	971	453	218	93	43	10	7	26.06
ZY NN	4	8	48	107	246	505	992	2043	3018	4049	5032	4984	3985	2999	1975	926	436	230	96	48	6	2	22.79

N= 512

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZYY 2NN	7	8	45	92	251	495	976	1919	2942	3919	4937	5074	4036	3003	1960	984	468	237	91	41	8	5	18.93
ZYY 3NN	4	8	41	83	237	477	969	1929	2869	3900	4955	5049	4028	3010	1966	983	494	234	93	39	5	3	20.42
ZYY 2N2N	5	11	52	120	275	505	1012	1979	2977	3959	4994	5012	4000	3011	1953	977	466	246	114	46	9	5	28.06
ZXX 2NN	5	12	71	126	298	530	1036	2029	2966	3948	4942	5062	4032	3054	2006	1003	508	262	109	43	7	4	29.60
ZXX 3NN	9	15	69	117	301	561	1030	1984	2951	3928	4929	5074	4103	3028	2070	1030	503	273	98	43	4	2	43.41
ZXX 2N2N	2	11	53	103	249	512	1010	2022	2994	3987	4980	5026	3948	2932	1960	989	494	247	101	55	14	7	16.61
ZXY 2NN	4	11	40	85	237	444	955	1932	2949	3984	4953	5050	3986	2946	1940	997	507	248	98	48	11	7	25.83
ZXY 3NN	1	8	42	88	230	477	981	1939	2942	3957	4934	5072	3991	2957	1998	999	499	251	95	52	6	2	21.45
ZXY 2N2N	3	8	58	101	232	469	992	1975	2987	4021	5050	4956	3967	2936	1935	968	468	255	111	53	18	9	26.95
ZYX 2NN	9	18	66	128	306	546	1071	1984	2953	3912	4933	5072	4079	3064	2102	1011	489	256	99	47	6	2	42.91
ZYX 3NN	9	15	66	122	290	557	1004	1965	2930	3932	4890	5114	4130	3084	2050	1013	502	248	93	42	6	2	28.62
ZYX 2N2N	3	13	61	128	267	538	1014	2023	3004	4003	4956	5048	4056	3039	1969	995	486	248	99	50	6	5	30.08
ZXA 2NN	3	9	52	104	252	533	997	2005	3017	3980	5013	4993	4008	3001	1957	1023	511	278	104	46	9	5	23.90
ZXA 3NN	6	11	58	109	267	501	1031	2006	2958	3959	4953	5054	4063	3033	2030	1031	505	253	106	55	8	3	12.26
ZXA 2N2N	0	4	41	98	236	484	1004	2020	3018	4049	5074	4927	3925	2943	1981	1001	498	245	106	62	17	8	19.38
ZAX 2NN	6	20	76	129	312	555	1054	1974	2946	3975	4923	5078	4090	3084	2044	1003	500	255	95	46	8	4	46.00
ZAX 3NN	12	19	76	128	311	582	1037	1967	2930	3956	4940	5061	4069	3106	2076	1030	496	253	95	39	5	3	49.81
ZAX 2N2N	7	14	68	130	265	517	1032	2012	2984	3965	4962	5043	4021	2985	1977	958	500	239	93	49	10	5	19.89
ZXB 2NN	3	8	46	88	212	457	956	1960	3001	4037	5035	4969	3981	2964	1952	1005	493	242	95	47	13	8	16.37
ZXB 3NN	2	6	35	84	234	468	943	1950	2977	3984	4973	5029	4027	2987	1967	955	498	249	99	50	5	3	17.38
ZXB 2N2N	0	2	41	95	233	477	997	1948	3033	4053	5069	4934	3940	2911	1945	981	486	232	105	58	15	8	29.28
ZBX 2NN	7	17	71	134	302	559	1047	1983	2962	3914	4912	5092	4102	3106	2076	1005	488	239	94	43	7	2	32.65
ZBX 3NN	8	18	74	135	311	561	1038	1957	2939	3924	4913	5088	4116	3109	2080	1017	479	246	91	42	5	3	41.15
ZBX 2N2N	6	13	75	125	284	541	1038	2006	3000	3980	4956	5050	4097	3033	1967	987	498	239	82	46	7	4	33.23
ZYA 2NN	9	21	64	115	289	535	1011	2023	2968	3965	4922	5089	4038	3045	2046	1023	515	249	106	48	7	2	31.06
ZYA 3NN	8	14	57	108	260	533	1019	1947	2939	3892	4943	5060	4067	3052	2028	1010	527	249	95	53	5	2	25.18
ZYA 2N2N	5	7	66	120	268	512	1026	2023	2994	4001	4994	5012	4032	3023	2011	1029	496	240	119	54	5	3	30.84
ZAY 2NN	5	8	42	92	240	464	979	1935	2940	3977	4974	5030	4018	2960	1942	1006	507	240	98	49	12	3	21.95
ZAY 3NN	4	10	39	97	229	471	980	1932	2929	3961	4973	5031	4009	2981	1958	1003	495	239	99	51	11	2	20.48
ZAY 2N2N	8	11	60	109	247	485	1005	1993	2977	4023	5020	4983	3969	2956	1926	952	474	240	98	62	18	9	24.56
ZX NN	4	10	52	107	247	500	1047	2042	3040	4020	5034	4976	3979	2972	2024	1001	517	252	108	51	10	7	15.35
ZY NN	4	10	52	97	222	496	1012	1994	3003	3962	4965	5045	4040	3005	1987	978	467	224	84	44	4	1	18.81

N=1024

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZYY 2NN	4	12	49	96	248	488	962	1978	3008	4039	5006	4997	4000	3013	2035	1000	512	225	100	55	13	7	20.92
ZYY 3NN	5	11	48	89	224	480	993	1992	2968	4008	5004	4998	4007	3030	2051	1009	502	258	99	44	12	7	12.74
ZYY 2N2N	5	9	49	100	238	489	967	1950	2946	4018	5024	4983	3973	2961	1950	1016	521	276	98	49	8	2	19.73
ZXX 2NN	4	11	57	103	265	500	984	1961	2939	3927	4922	5083	4084	3031	2013	995	475	240	84	50	16	8	20.83
ZXX 3NN	5	9	61	111	265	505	992	1970	2893	3904	4924	5079	4051	3009	2026	997	489	226	96	48	9	5	18.89
ZXX 2N2N	3	4	47	99	252	486	1000	2011	2996	3978	4953	5054	4028	2978	1990	1054	544	277	107	49	7	3	20.91
ZXY 2NN	4	7	56	94	238	498	966	1978	3021	3994	5012	4995	4006	3029	2014	1013	498	241	98	60	17	9	21.57
ZXY 3NN	6	8	44	94	242	497	994	2025	2970	4008	4998	5004	4006	3006	2023	1026	507	240	107	44	15	8	21.21
ZXY 2N2N	3	7	39	101	256	484	1011	1987	3020	4039	5011	4993	3944	2948	2007	1000	522	283	118	55	12	4	26.28
ZYX 2NN	7	10	50	101	239	513	980	1897	2940	3917	4905	5097	4069	3028	1991	986	479	253	98	46	8	4	23.45
ZYX 3NN	8	12	58	109	261	502	981	1934	2875	3929	4925	5083	4075	3037	2013	985	490	237	97	47	11	4	17.80
ZYX 2N2N	9	12	48	87	254	486	970	1942	2911	3922	4930	5073	3996	2974	1946	986	526	273	106	45	4	3	32.64
ZXA 2NN	2	10	57	96	243	476	964	1979	2975	3999	4979	5027	4043	3068	2004	1006	504	265	102	54	20	11	28.18
ZXA 3NN	5	10	56	103	249	497	986	1963	2958	3948	4970	5032	4018	3054	2001	1001	491	231	99	53	11	9	14.99
ZXA 2N2N	4	7	46	99	255	490	990	2004	2976	4018	4971	5030	4002	2970	2010	1028	547	300	120	59	8	5	24.05
ZAX 2NN	7	13	57	110	262	521	999	1915	2909	3886	4909	5099	4103	3036	2022	993	471	226	86	49	10	2	25.21
ZAX 3NN	7	11	64	116	264	517	1004	1927	2871	3888	4901	5102	4065	3027	2014	987	482	223	85	47	6	2	25.03
ZAX 2N2N	2	10	52	112	248	522	980	1970	2963	3934	4946	5055	4052	3035	2023	1010	513	266	99	42	4	1	21.63
ZXB 2NN	7	11	47	92	248	496	956	1997	3003	4022	5006	5002	4026	3016	1994	1006	508	284	113	57	18	13	28.54
ZXB 3NN	4	10	42	95	233	486	993	2024	2991	4016	4978	5028	4022	3000	1969	999	517	254	100	55	14	7	13.54
ZXB 2N2N	4	6	51	99	231	507	1002	2019	2990	4064	5043	4958	3950	2979	2025	1003	530	291	125	68	13	6	29.44
ZBX 2NN	7	10	60	108	266	504	1006	1916	2926	3887	4897	5105	4081	3032	2013	969	470	241	89	45	7	3	23.83
ZBX 3NN	7	12	64	103	262	529	995	1942	2858	3880	4897	5104	4067	3044	2017	978	463	242	95	34	6	3	37.05
ZBX 2N2N	4	10	45	104	264	509	958	1932	2922	3888	4942	5062	4038	2976	1972	984	501	269	91	40	6	1	28.96
ZYA 2NN	9	12	49	98	248	487	962	1946	2937	3960	4962	5045	4041	2989	1990	991	505	246	96	53	9	3	12.65
ZYA 3NN	7	9	50	101	253	475	955	1955	2899	3936	4969	5032	4035	3041	1989	1014	480	234	90	46	12	4	21.78
ZYA 2N2N	10	12	49	107	235	480	944	1978	2955	3929	4999	5005	4007	2995	1988	992	509	267	103	52	9	4	24.02
ZAY 2NN	3	12	45	96	243	490	1018	1977	2999	3997	4988	5012	3993	3055	2024	996	500	241	102	51	14	9	19.72
ZAY 3NN	5	8	43	91	246	468	1007	2009	3004	3993	5009	4992	3977	3020	2052	1005	501	243	106	44	14	8	22.33
ZAY 2N2N	2	6	34	86	246	516	959	1991	2968	3984	5068	4935	3921	2951	1980	998	502	270	115	52	7	1	34.18
ZX NN	3	10	51	92	239	479	981	1949	2937	3944	4947	5063	4019	3013	2043	1028	519	246	108	52	10	5	11.67
ZY NN	4	8	55	99	243	473	980	1981	2992	3978	4962	5046	4032	3030	2023	1004	526	253	114	57	13	6	11.79

N= 2

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
TYY 2NN	7	13	56	95	250	474	983	1987	3017	4041	5095	5012	3974	2972	1992	992	514	250	97	49	7	5	12.61
TYY 3NN	7	16	56	98	240	460	964	1968	2990	4035	5094	4993	3965	2954	1986	1008	501	269	103	58	13	9	21.60
TYY 2N2N	6	10	42	88	268	504	975	1971	2897	3892	5050	5077	3969	2932	1926	981	511	234	80	42	8	4	40.65
TXX 2NN	60	70	110	126	208	364	1477	2455	3147	3866	4663	5676	4926	3871	1518	585	276	148	54	34	7	3	3954.63
TXX 3NN	65	66	90	113	260	1010	1550	2224	2874	3428	4015	6165	5507	3529	1219	465	213	119	45	27	10	5	5640.43
TXX 2N2N	1	3	10	26	118	228	723	2562	3663	4253	5229	5155	4259	3684	2598	739	259	117	43	16	4	2	2218.36
TXY 2NN	267	391	800	1081	1514	1942	2505	3251	3846	4435	5243	4862	3983	2983	1937	1029	623	440	296	217	98	74	23431.88
TXY 3NN	531	671	1067	1294	1644	1989	2372	3020	3602	4312	5133	4934	4014	3025	2012	1108	729	530	379	305	176	148	68000.43
TXY 2N2N	65	103	244	390	742	1127	1727	2737	3628	4498	5421	4706	3697	2703	1592	582	215	89	52	39	9	5	2018.21
TYX 2NN	6	10	48	69	121	189	506	1497	2363	3142	4127	5981	4839	3558	2229	1028	515	263	112	60	18	7	596.35
TYX 3NN	18	27	60	88	119	272	744	1509	2116	2758	3740	6344	5005	3406	1890	693	277	123	54	28	12	9	1437.74
TYX 2N2N	8	13	38	51	91	210	577	1621	2653	3677	4654	5455	4519	3713	2781	1710	1065	692	388	256	90	58	1790.03
TYB 2NN	3	5	31	57	145	295	639	1496	2522	3719	5092	5057	3727	2592	1530	646	306	122	51	24	5	2	517.60
TYB 3NN	4	6	17	37	84	189	432	1218	2230	3559	5010	5037	3555	2306	1253	449	170	81	33	19	5	4	1295.28
TYB 2N2N	10	21	73	129	293	573	1051	1980	2941	3929	4977	5098	4008	2993	2020	1043	540	281	123	57	13	6	40.74
TBY 2NN	26	49	165	285	592	986	1658	2659	3509	4332	5096	4964	4183	3437	2634	1601	1022	604	293	165	42	21	1687.33
TBY 3NN	51	79	263	432	787	1244	1861	2862	3671	4360	5079	4950	4260	3575	2821	1866	1247	829	461	310	93	51	4439.44
TBY 2N2N	12	21	79	141	306	586	1056	2025	2990	3961	5065	5036	3981	2960	1984	1067	555	284	126	64	15	8	54.68
TXA 2NN	83	108	300	456	850	1637	3485	4860	5578	6087	6663	3505	3075	1337	344	34	5	1	1	1	0	0	10385.28
TXA 3NN	85	101	217	374	1107	2072	2934	4046	4805	5389	5757	4277	3569	996	134	13	1	0	0	0	0	0	11251.16
TXA 2N2N	97	143	502	791	1377	2254	3768	5612	6044	6287	6587	3472	3039	2526	872	218	30	11	0	0	0	0	14529.19
TAX 2NN	58	64	83	100	122	145	560	1689	2033	2233	2461	7622	7362	6973	5324	3150	2308	1698	946	656	196	115	17762.45
TAX 3NN	72	76	94	108	138	557	1323	1843	2041	2186	2342	7703	7534	7071	5193	2864	2176	1740	1214	886	403	287	33169.67
TAX 2N2N	0	0	1	2	16	38	266	896	2476	3019	3421	6631	6358	6105	5686	3794	2185	1368	785	503	162	99	14682.17
TXB 2NN	174	276	798	1144	1799	2425	3141	3980	4530	5013	5662	4487	3432	2067	904	359	198·	134	73	44	13	9	18992.64
TXB 3NN	389	551	1085	1420	1920	2356	2844	3440	3885	4436	5319	4732	3437	1963	855	341	223	149	108	69	21	14	44856.03
TXB 2N2N	98	156	460	749	1332	1958	2846	4075	4862	5468	6017	4041	3401	2389	1113	255	61	23	6	0	0	0	8597.88
TBX 2NN	28	46	80	99	114	135	383	1666	2124	2479	3006	7059	6307	5254	3799	2086	1091	595	248	126	28	20	3685.70
TBX 3NN	68	82	105	110	116	291	1147	1826	2036	2243	2648	7386	6670	5489	3846	1875	872	432	154	86	22	16	5287.05
TBX 2N2N	1	1	3	6	29	68	265	1093	2294	3327	4030	6037	5514	4909	4117	2876	1954	1279	738	473	154	86	8330.39
TYA 2NN	3	7	30	50	111	237	539	1375	2358	3449	4789	5362	3984	2695	1573	687	334	166	65	42	12	4	632.48
TYA 3NN	5	6	25	41	84	191	468	1141	2015	3130	4515	5552	3853	2361	1129	367	137	64	27	21	6	4	1762.52
TYA 2N2N	26	33	68	111	236	470	959	1916	2909	3873	4888	5194	4179	3236	2208	1256	748	488	270	169	63	48	689.00
TAY 2NN	356	449	799	1007	1364	1747	2277	3062	3723	4414	5114	4948	4211	3468	2670	1733	1172	815	533	431	241	180	37550.62
TAY 3NN	611	711	1037	1228	1558	1928	2419	3183	3811	4447	5055	4981	4299	3594	2885	2005	1460	1060	736	588	375	316	99925.75
TAY 2N2N	49	74	177	265	514	811	1299	2218	3162	4083	5110	4967	3922	2906	1889	916	467	229	104	66	26	16	729.18
TD 2N	20	36	122	228	462	802	1314	2233	3162	4153	5318	4922	3779	2712	1693	800	380	199	75	35	7	3	338.03
TX NN	64	68	89	136	232	429	741	1749	3626	4237	5428	5300	4128	3532	1741	747	427	224	129	84	67	65	3412.54
TY NN	8	14	45	95	256	503	968	2006	3058	4084	5148	5017	3941	2989	2011	997	496	236	103	49	7	6	21.66

N= 4

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
TYY 2NN	3	8	54	114	251	478	986	2029	3049	4009	5044	5017	3997	3031	2076	1010	518	236	103	53	10	7	26.56
TYY 3NN	5	8	44	100	249	473	970	2008	2985	3987	4966	5082	4042	3014	2011	1019	486	248	83	39	9	6	18.24
TYY 2N2N	6	10	68	115	266	518	985	2012	3019	4037	5086	5016	3984	2958	2014	1051	508	259	119	59	10	5	25.92
TXX 2NN	2	4	26	64	263	451	1439	2372	3136	3873	4581	5563	4945	4048	1978	414	142	62	22	14	4	4	2827.04
TXX 3NN	5	9	109	161	336	872	1401	2166	2948	3558	4106	5972	5363	4054	1695	248	83	31	10	7	2	0	3820.12
TXX 2N2N	0	1	9	16	72	250	941	2558	3528	4228	5087	5063	4212	3526	2575	983	264	81	21	8	2	1	1473.82
TXY 2NN	219	294	569	739	1086	1450	1938	2730	3539	4319	5211	4876	3938	2894	1823	780	332	171	89	56	19	15	12384.00
TXY 3NN	316	379	600	742	1021	1301	1768	2608	3368	4178	5082	4972	3979	2937	1878	865	398	215	105	68	29	16	21334.23
TXY 2N2N	65	95	232	357	652	1009	1611	2572	3465	4376	5351	4731	3685	2644	1672	681	236	77	17	6	0	0	1589.81
TYX 2NN	0	0	3	18	81	230	601	1602	2545	3448	4368	5706	4724	3608	2499	1207	602	303	117	58	17	12	384.98
TYX 3NN	1	2	15	39	138	326	738	1653	2529	3283	4142	5890	4913	3695	2316	898	344	114	31	10	2	1	647.70
TYX 2N2N	2	2	5	18	96	243	653	1671	2755	3759	4799	5291	4344	3482	2607	1662	1045	650	368	249	94	62	1665.40
TYB 2NN	1	4	15	35	121	275	620	1509	2619	3766	5118	4990	3789	2624	1568	658	270	126	39	21	5	5	486.52
TYB 3NN	0	1	7	12	45	148	425	1240	2347	3603	4951	5087	3647	2377	1290	429	136	48	12	6	0	0	1206.17
TYB 2N2N	9	16	69	146	287	556	1034	2044	3040	4057	5086	4981	3969	3043	2042	1036	568	306	135	71	15	7	51.24
TBY 2NN	36	54	188	304	591	950	1551	2549	3387	4185	5061	4986	4238	3417	2561	1612	973	607	294	174	53	30	1668.58
TBY 3NN	65	101	300	453	770	1189	1814	2734	3518	4269	4992	5032	4318	3608	2786	1828	1213	813	454	291	96	65	4611.62
TBY 2N2N	5	16	72	125	301	547	1024	2041	2976	3980	5070	4990	3960	2944	2015	1057	549	301	131	59	16	6	52.52
TXA 2NN	56	99	412	788	1543	2643	3506	4097	4530	5033	5605	4444	3597	1729	535	45	3	0	0	0	0	0	13328.21
TXA 3NN	36	86	411	717	1514	1967	2431	3020	3626	4210	4935	5086	3433	1671	350	14	0	0	0	0	0	0	9985.20
TXA 2N2N	210	301	660	976	1763	2846	3844	4288	4584	5009	5580	4453	3652	2613	1240	292	76	17	1	0	0	0	21868.64
TAX 2NN	0	2	11	20	131	301	643	1951	2728	3186	3438	6579	6427	6292	5827	3798	2236	1611	1172	887	442	320	35150.56
TAX 3NN	5	11	45	151	306	521	1195	2290	2892	3253	3438	6568	6464	6329	5789	3558	1813	1259	965	806	475	379	40251.70
TAX 2N2N	0	0	1	3	13	72	309	1254	2599	3659	4446	5601	5007	4584	4330	3908	2850	1768	950	648	291	196	21163.57
TXB 2NN	340	468	925	1257	1706	2131	2681	3304	3815	4521	5444	4644	3457	2110	866	161	42	19	9	5	1	0	33611.78
TXB 3NN	413	531	879	1064	1340	1630	1978	2532	3263	4154	5260	4782	3384	1949	742	129	29	13	9	5	2	2	40319.98
TXB 2N2N	196	292	706	1011	1592	2139	2631	3613	4182	4806	5576	4469	3524	2450	1302	342	87	20	2	0	0	0	16354.42
TBX 2NN	0	0	8	19	88	284	620	1747	2764	3307	3724	6309	5862	5326	4409	2875	1710	995	467	264	60	33	5687.64
TBX 3NN	4	8	31	87	307	469	958	2212	2934	3322	3564	6448	6119	5566	4620	2902	1611	808	309	137	24	8	6092.36
TBX 2N2N	0	0	1	2	15	87	360	1317	2449	3524	4525	5520	4760	4187	3622	2832	2139	1599	1017	703	304	206	17041.89
TYA 2NN	1	1	8	21	78	218	550	1411	2432	3605	4873	5239	3925	2748	1610	663	297	140	49	27	7	4	560.09
TYA 3NN	0	0	7	18	50	136	415	1180	2147	3315	4670	5368	3861	2483	1239	360	108	28	5	2	0	0	1453.32
TYA 2N2N	2	6	48	92	239	464	935	1913	2939	3954	4981	5090	4058	3129	2224	1226	672	406	214	139	40	26	245.14
TAY 2NN	173	230	438	599	914	1273	1835	2739	3536	4299	5058	4979	4223	3394	2523	1529	917	552	277	180	70	47	8429.52
TAY 3NN	268	358	605	769	1091	1490	2057	2910	3607	4341	5018	5015	4299	3590	2766	1820	1200	791	461	298	132	96	20259.86
TAY 2N2N	28	67	127	197	379	673	1199	2157	3106	4093	5176	4883	3850	2823	1936	975	471	220	79	33	6	3	238.07
TD 2N	33	56	154	234	432	732	1230	2199	3128	4107	5189	4974	3877	2830	1776	822	374	177	58	26	7	3	396.85
TX NN	1	3	17	36	124	233	748	2599	3661	4245	5195	5169	4192	3601	2605	790	269	138	42	18	5	2	2086.30
TY NN	6	11	42	92	253	491	1006	2006	3018	4061	5082	5034	3988	3035	2078	1041	515	247	110	48	9	4	19.87

N= 8

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
TYY 2NN	1	7	52	99	246	516	985	1957	3004	3969	5001	5044	3986	3010	1986	1035	496	220	87	48	5	3	29.83
TYY 3NN	7	11	47	100	242	500	990	1995	2996	4031	5069	4958	3969	3004	1990	1005	501	253	92	44	8	4	8.52
TYY 2N2N	5	12	51	94	250	482	969	1935	2922	3924	5013	5052	3973	3036	2073	1002	489	251	91	37	8	3	26.47
TXX 2NN	3	9	52	100	339	663	1314	2224	3046	3830	4614	5422	4661	3765	2361	600	97	20	6	5	2	1	1413.70
TXX 3NN	17	32	121	223	415	755	1269	2116	2933	3732	4492	5540	4707	3829	2401	385	30	6	3	2	1	1	2118.76
TXX 2N2N	0	1	9	28	145	452	1118	2201	3198	4021	4946	5119	4161	3283	2257	1057	418	143	25	6	0	0	345.37
TXY 2NN	151	198	348	473	728	1038	1532	2446	3251	4128	5076	4981	3916	2885	1835	742	288	119	43	19	4	4	5300.77
TXY 3NN	138	174	322	447	676	971	1464	2330	3231	4091	5064	4971	3959	2891	1853	832	347	151	55	25	8	5	4293.36
TXY 2N2N	38	62	190	274	516	830	1369	2267	3203	4133	5151	4912	3864	2843	1770	701	251	99	18	3	0	0	756.84
TYX 2NN	0	0	17	42	134	290	713	1659	2639	3654	4598	5445	4477	3456	2418	1226	581	284	119	58	10	7	214.52
TYX 3NN	5	10	38	73	175	381	821	1727	2667	3593	4512	5517	4516	3523	2344	1052	399	147	37	16	3	0	201.32
TYX 2N2N	0	0	9	27	90	272	710	1701	2763	3835	4761	5308	4294	3301	2370	1346	823	514	263	169	69	47	780.15
TYB 2NN	0	0	12	28	97	244	602	1472	2572	3738	5013	5054	3769	2600	1540	596	221	85	22	8	1	1	585.59
TYB 3NN	0	0	7	11	44	140	409	1283	2372	3648	5005	5023	3664	2366	1236	429	127	36	8	5	1	1	1215.53
TYB 2N2N	10	16	57	103	248	498	1010	1984	2956	3951	4967	5070	3995	3024	2043	1041	499	253	100	55	11	4	17.44
TBY 2NN	31	52	196	317	563	921	1493	2477	3355	4206	5015	5030	4190	3368	2494	1530	936	566	297	179	52	30	1475.35
TBY 3NN	74	125	296	452	787	1169	1767	2731	3520	4333	5044	4974	4271	3502	2717	1793	1168	779	474	310	106	72	4981.82
TBY 2N2N	3	14	51	108	252	505	981	1968	2907	3930	4964	5074	3998	3027	2009	1015	496	254	98	52	9	5	21.78
TXA 2NN	111	205	683	993	1491	1767	2026	2577	3263	4103	5108	4956	3627	2215	851	101	7	0	0	0	0	0	12501.07
TXA 3NN	118	211	508	653	816	998	1312	1981	2775	3780	4982	5039	3626	2021	607	27	0	0	0	0	0	0	8035.12
TXA 2N2N	152	238	705	1109	1579	1839	2117	2769	3524	4366	5321	4719	3730	2690	1549	502	143	36	3	1	0	0	14022.07
TAX 2NN	3	6	27	55	186	444	968	1969	2919	3758	4476	5552	4971	4577	4316	3845	2754	1505	720	534	342	297	27219.90
TAX 3NN	13	23	103	188	411	783	1310	2262	3087	3868	4532	5487	4924	4588	4323	3877	2711	1358	525	337	197	159	16323.10
TAX 2N2N	0	0	1	4	37	139	517	1536	2610	3610	4621	5410	4463	3631	2833	2101	1847	1557	1093	709	245	183	14949.11
TXB 2NN	332	429	693	845	1081	1351	1733	2423	3255	4151	5245	4834	3523	2214	1011	158	26	6	2	2	1	1	25700.90
TXB 3NN	245	291	427	505	678	899	1255	2003	2879	3939	5180	4842	3439	2039	816	108	10	3	3	2	1	1	13538.33
TXB 2N2N	265	348	675	890	1171	1474	1975	2751	1582	4416	5344	4694	3684	2612	1475	482	148	30	3	0	0	0	17024.67
TBX 2NN	2	4	19	53	160	376	876	1853	2804	3677	4434	5585	4940	4395	3864	2965	2147	1446	767	451	135	78	8501.71
TBX 3NN	9	18	82	152	334	672	1250	2115	2967	3774	4526	5492	4936	4505	4008	3145	2219	1400	685	363	73	38	8249.09
TBX 2N2N	0	0	1	4	41	142	480	1456	2542	3596	4566	5454	4497	3609	2813	1991	1525	1193	878	680	342	248	16683.11
TYA 2NN	0	0	6	22	85	223	528	1411	2459	3613	4868	5207	3925	2689	1591	591	225	99	29	10	2	0	642.50
TYA 3NN	1	1	9	21	55	163	415	1201	2250	3463	4772	5254	3811	2445	1215	351	93	18	7	1	0	0	1382.39
TYA 2N2N	5	6	31	79	194	443	933	1900	2957	3932	4916	5123	4118	3085	2098	1097	574	304	141	88	27	14	68.87
TAY 2NN	102	147	323	437	724	1095	1676	2565	3409	4204	5063	4964	4179	3352	2465	1465	875	527	262	151	46	33	3633.19
TAY 3NN	175	218	443	594	959	1340	1922	2795	3566	4339	5005	5005	4273	3482	2679	1741	1130	740	424	265	100	63	9707.18
TAY 2N2N	8	18	75	147	304	565	1098	2044	2994	3967	4989	5045	3997	2973	1965	925	433	195	71	29	4	2	51.67
TD 2N	22	38	120	200	381	649	1147	2115	3081	4052	5088	5007	3992	2904	1912	874	376	172	59	29	4	1	200.50
TX NN	0	3	13	26	79	267	994	2525	3468	4166	5005	5134	4280	3575	2551	953	277	80	24	9	3	2	1352.18
TY NN	3	9	59	113	245	487	1011	1968	2965	3972	4995	5098	4037	3014	2008	974	450	239	88	46	12	5	22.00

N= 16

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES										RIGHT-TAIL CUMULATIVE PROBABILITIES												
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
TYY 2NN	5	7	44	87	237	543	1005	1987	3014	4052	5112	4928	3918	2940	1960	961	440	199	75	40	9	4	35.96
TYY 3NN	5	9	44	91	257	490	1007	2010	3035	4089	5050	4978	3908	2935	1987	979	458	212	80	39	6	2	25.39
TYY 2N2N	0	9	40	91	237	486	1019	2003	2991	4001	5065	4976	3899	2920	1947	933	436	209	85	42	8	5	28.80
TXX 2NN	10	16	80	160	364	670	1203	2120	2997	3889	4805	5229	4275	3313	2215	971	260	46	2	0	0	0	455.41
TXX 3NN	24	38	125	200	400	658	1135	2053	2893	3793	4712	5309	4370	3315	2203	947	187	27	2	0	0	0	678.87
TXX 2N2N	0	0	24	72	221	531	1116	2170	3142	4082	5121	4936	3941	3036	2102	1047	481	216	67	21	0	0	99.24
TXY 2NN	87	107	237	350	613	896	1375	2350	3277	4223	5148	4877	3876	2829	1793	791	305	99	30	14	1	0	1883.63
TXY 3NN	56	84	203	304	537	834	1338	2240	3204	4170	5139	4883	3886	2902	1850	843	340	135	42	22	3	1	969.67
TXY 2N2N	36	54	152	237	480	765	1300	2338	3304	4255	5156	4895	3856	2795	1783	773	336	129	35	15	2	0	503.52
TYX 2NN	0	2	23	49	144	371	834	1779	2762	3748	4769	5268	4274	3318	2305	1205	630	317	105	49	10	8	125.68
TYX 3NN	2	4	43	86	207	421	881	1797	2778	3724	4682	5345	4318	3317	2256	1127	508	213	60	21	2	0	109.28
TYX 2N2N	0	0	9	25	110	318	788	1863	2923	3927	4972	5077	4058	3140	2200	1210	709	433	208	132	37	21	319.03
TYB 2NN	0	0	7	24	83	223	587	1536	2635	3850	5105	4937	3720	2477	1462	521	183	66	16	7	2	0	675.11
TYB 3NN	0	0	3	10	48	125	437	1290	2415	3701	5073	4946	3587	2313	1200	352	101	31	3	2	0	0	1285.72
TYB 2N2N	2	6	44	89	249	505	1001	2009	3049	4089	5131	4888	3891	2904	1931	930	443	219	79	39	10	5	17.55
TBY 2NN	32	52	173	294	550	910	1488	2484	3384	4277	5107	4912	4081	3298	2448	1514	913	541	250	144	51	32	128[illegible].04
TBY 3NN	70	102	292	432	754	1158	1746	2744	3571	4317	5076	4944	4161	3428	2670	1754	1157	771	428	256	103	62	4235.80
TBY 2N2N	6	16	53	105	246	514	1022	1992	2987	3969	5060	4970	3970	2940	1951	954	486	247	100	55	18	9	25.47
TXA 2NN	178	242	374	443	593	860	1304	2159	3040	4047	5143	4891	3648	2301	1049	200	29	1	0	0	0	0	7869.14
TXA 3NN	71	81	144	203	341	552	951	1735	2630	3734	4997	5034	3597	2098	816	105	7	1	0	0	0	0	2447.31
TXA 2N2N	201	272	416	513	784	1111	1682	2601	3499	4390	5303	4720	3747	2713	1637	606	214	60	12	1	0	0	9256.30
TAX 2NN	4	10	50	109	286	574	1101	2100	2972	3780	4543	5476	4677	3903	3067	2293	1995	1750	1218	749	163	111	15009.14
TAX 3NN	24	43	141	254	484	835	1421	2342	3138	3850	4528	5487	4789	4060	3261	2378	2012	1775	1257	739	115	63	15796.73
TAX 2N2N	0	0	2	10	79	232	664	1673	2806	3859	4872	5159	4236	3369	2498	1602	1064	766	502	418	286	210	9992.96
TXB 2NN	176	215	340	424	637	896	1367	2210	3145	4127	5297	4759	3529	2258	1070	229	31	6	0	0	0	0	7158.99
TXB 3NN	72	88	156	216	377	572	925	1789	2778	3924	5205	4814	3432	2087	909	153	22	3	0	0	0	0	2192.20
TXB 2N2N	211	254	433	557	805	1182	1697	2657	3513	4356	5321	4709	3714	2632	1580	576	203	62	6	1	0	0	9784.08
TBX 2NN	1	6	43	85	228	477	1045	2027	2953	3784	4625	5396	4646	3909	3148	2312	1834	1466	1005	704	284	179	13925.11
TBX 3NN	20	35	126	207	412	772	1377	2275	3102	3879	4577	5433	4737	4058	3297	2445	1933	1521	1069	759	255	137	14188.01
TBX 2N2N	0	0	2	11	52	196	615	1652	2728	3816	4844	5183	4232	3382	2492	1633	1103	777	525	411	250	200	8842.69
TYA 2NN	0	1	8	18	76	208	577	1498	2555	3732	4954	5085	3821	2602	1490	562	195	67	16	8	0	0	652.30
TYA 3NN	0	0	6	18	54	150	434	1265	2340	3494	4885	5132	3748	2383	1196	362	96	22	8	0	0	0	1325.56
TYA 2N2N	0	5	35	70	206	437	956	2012	3034	4024	5092	4934	3984	2972	1951	1006	541	273	103	44	18	10	45.65
TAY 2NN	71	106	269	409	681	1009	1606	2595	3449	4309	5103	4916	4129	3264	2406	1439	859	510	229	120	39	24	2280.52
TAY 3NN	126	172	387	540	863	1256	1859	2814	3616	4350	5086	4936	4203	3446	2633	1729	1125	739	395	242	84	59	6482.88
TAY 2N2N	13	25	72	138	304	586	1093	2105	3100	4069	5079	4950	3957	2925	1930	942	468	216	97	46	12	8	48.35
TD 2N	17	29	96	161	366	623	1130	2120	3108	4077	5045	5028	3943	2924	1880	864	397	185	66	22	3	1	114.41
TX NN	1	1	8	32	170	472	1124	2274	3261	4119	5056	5007	4079	3235	2294	1106	449	154	31	5	0	0	346.97
TY NN	4	10	47	98	254	506	1017	2040	2991	4007	5076	5000	3941	2930	1966	984	480	225	73	36	11	6	16.84

N= 32

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
TYY 2NN	5	9	43	91	226	474	996	1961	2936	3944	4982	5041	4029	3004	1994	1014	513	259	103	57	13	10	14.23
TYY 3NN	6	8	43	89	240	482	1005	1953	2981	3973	4977	5044	4066	3006	2003	1017	512	271	113	51	12	6	16.83
TYY 2N2N	5	6	48	91	240	450	977	1987	2963	3962	4977	5064	3950	2998	2003	983	493	262	108	55	12	6	26.88
TXX 2NN	2	8	53	115	291	555	1041	1971	2862	3862	4834	5191	4224	3163	2063	977	393	170	36	6	0	0	118.43
TXX 3NN	3	12	77	144	322	595	1080	1979	2882	3794	4720	5300	4284	3220	2105	938	383	128	21	4	0	0	195.51
TXX 2N2N	0	2	34	81	222	492	978	1990	2982	3996	4973	5056	4037	3051	2005	991	496	238	93	46	6	0	21.30
TXY 2NN	37	59	146	238	434	682	1190	2153	3101	4059	5075	4949	3941	2940	1887	836	375	164	50	18	3	1	422.35
TXY 3NN	31	44	117	218	424	705	1211	2141	3101	4027	5002	5012	3965	2977	1950	898	411	193	71	27	2	2	293.05
TXY 2N2N	20	34	113	163	367	627	1181	2148	3118	4048	5070	4964	3910	2849	1848	841	367	164	54	28	2	1	180.63
TYX 2NN	0	0	17	48	152	354	831	1779	2750	3745	4756	5271	4232	3156	2155	1151	600	337	160	82	19	8	113.49
TYX 3NN	1	3	36	72	186	414	878	1843	2775	3731	4753	5265	4301	3221	2155	1072	537	280	96	46	6	2	63.29
TYX 2N2N	1	1	16	30	134	329	806	1814	2818	3836	4869	5160	4133	3092	2137	1185	663	389	194	124	31	21	231.12
TYB 2NN	1	2	7	2	79	200	564	1486	2595	3747	5013	5027	3738	2534	1493	591	209	86	26	10	0	0	652.00
TYB 3NN	0	1	3	7	35	115	385	1261	2329	3660	4993	5023	3586	2346	1251	430	131	33	10	4	0	0	1287.00
TYB 2N2N	3	6	42	93	251	513	976	1963	2965	3997	5034	4985	3962	2963	1961	969	517	257	89	46	8	5	17.01
TBY 2NN	34	52	170	284	544	891	1460	2423	3308	4096	4902	5116	4241	3364	2460	1524	959	556	316	216	60	35	1514.30
TBY 3NN	74	115	283	433	750	1124	1713	2674	3490	4253	4935	5073	4351	3553	2732	1798	1191	787	471	329	135	95	5353.18
TBY 2N2N	5	10	42	86	231	468	971	1986	2958	3944	4969	5047	4057	3013	2061	1004	511	273	104	51	13	5	17.65
TXA 2NN	30	42	110	177	341	547	968	1851	2819	3863	5037	4991	3709	2330	1149	285	58	11	1	0	0	0	1033.26
TXA 3NN	10	15	43	84	174	374	723	1552	2552	3665	4997	5016	3593	2209	945	162	19	2	0	0	0	0	1393.58
TXA 2N2N	61	89	214	311	563	838	1347	2272	3213	4183	5167	4855	3866	2764	1699	676	261	98	27	10	0	0	1196.82
TAX 2NN	2	3	41	105	203	599	1138	2127	3018	3832	4628	5384	4551	3757	2933	1990	1391	989	655	502	313	229	12620.33
TAX 3NN	16	31	146	254	515	898	1492	2362	3189	3888	4630	5380	4673	3939	3154	2187	1577	1134	753	560	323	247	15465.29
TAX 2N2N	0	0	4	15	80	241	676	1718	2751	3767	4774	5249	4239	3298	2335	1370	829	559	302	226	110	84	1724.87
TXB 2NN	51	69	137	191	365	566	1005	1881	2897	4010	5145	4906	3640	2358	1201	318	78	13	3	0	0	0	1210.85
TXB 3NN	15	20	52	85	186	367	725	1623	2646	3826	5091	4930	3547	2172	1009	227	45	5	1	1	0	0	1183.12
TXB 2N2N	97	117	239	343	591	889	1429	2306	3267	4176	5173	4848	3780	2745	1643	673	257	98	30	9	0	0	2224.52
TBX 2NN	0	4	36	100	249	542	1099	2099	2977	3801	4627	5388	4564	3787	2936	1963	1406	1035	688	534	309	237	13268.09
TBX 3NN	13	27	132	217	459	840	1405	2360	3123	3873	4620	5389	4653	3923	3127	2205	1598	1152	790	595	330	255	16314.65
TBX 2N2N	0	0	2	17	77	229	641	1656	2690	3746	4774	5245	4239	3299	2384	1433	882	583	364	255	132	92	2243.79
TYA 2NN	0	0	4	15	71	193	533	1443	2511	3680	4935	5092	3787	2596	1491	581	219	89	21	5	1	0	692.88
TYA 3NN	0	0	2	12	46	115	375	1189	2295	3530	4906	5112	3722	2349	1215	391	110	31	4	1	0	0	1388.97
TYA 2N2N	2	4	30	79	217	444	943	1980	2972	3947	4966	5053	4011	3011	1988	1014	539	291	109	53	9	4	25.42
TAY 2NN	50	82	227	348	604	953	1516	2473	3349	4173	4993	5018	4219	3372	2482	1500	901	523	267	174	53	33	1773.23
TAY 3NN	108	150	330	508	849	1208	1781	2703	3517	4259	4979	5032	4324	3573	2736	1791	1159	762	425	291	123	81	6217.65
TAY 2N2N	7	14	51	103	248	529	1006	2020	3002	3972	4978	5042	4021	2979	1991	989	482	236	97	48	8	5	12.23
TD 2N	8	18	87	148	317	578	1069	2069	3077	4099	5114	4947	3899	2921	1932	903	428	188	61	25	4	0	59.16
TX NN	0	0	18	58	201	478	1056	2060	3071	4026	5021	5017	4071	3101	2083	1048	491	207	76	24	1	1	79.06
TY NN	2	9	45	100	249	467	950	1949	2999	3945	5028	5034	3991	2994	2017	1047	524	261	112	50	5	3	25.73

BLANK PAGE

N= 64

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES										RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
TYY 2NN	6	14	58	111	277	529	1042	1995	2962	3949	4975	5039	4024	2998	1945	962	460	235	88	42	7	2	17.62
TYY 3NN	10	12	66	119	267	532	1030	2043	2978	3946	4989	5024	4014	2997	1968	974	469	229	101	43	11	2	32.30
TYY 2N2N	5	9	51	105	254	504	1006	2005	3028	3990	5000	5020	4014	3002	1961	960	479	233	96	40	11	4	11.11
TXX 2NN	7	11	68	132	314	565	1098	2061	3017	3966	4958	5061	4047	3063	1989	956	432	193	46	21	1	0	66.60
TXX 3NN	12	17	74	159	347	623	1108	2063	3014	3964	4917	5093	4147	3135	2037	957	405	170	40	14	0	0	130.96
TXX 2N2N	2	8	33	80	223	505	1027	2088	3131	4089	5093	4929	3952	2965	1977	952	460	228	84	37	4	1	29.74
TXY 2NN	15	30	110	186	398	708	1198	2165	3138	4080	5070	4944	3956	2901	1868	828	376	155	42	10	3	1	201.10
TXY 3NN	25	38	116	205	369	669	1162	2134	3129	4057	5021	5001	3928	2932	1907	898	394	162	59	15	3	1	224.87
TXY 2N2N	11	22	86	154	325	601	1125	2154	3166	4139	5112	4905	3889	2825	1836	865	381	147	51	20	2	0	97.26
TYX 2NN	2	4	32	69	219	470	930	1915	2906	3872	4852	5164	4114	3104	2156	1071	566	299	123	54	17	11	48.38
TYX 3NN	5	9	49	97	252	517	1001	1921	2899	3821	4838	5174	4125	3130	2102	1056	522	265	100	46	9	2	26.87
TYX 2N2N	1	2	23	56	188	420	892	1904	2967	4000	4960	5064	4022	3038	2116	1097	601	325	163	77	19	13	92.45
TYB 2NN	0	1	13	25	98	247	595	1518	2623	3805	4981	5045	3801	2575	1509	539	212	70	16	7	1	0	608.23
TYB 3NN	0	0	3	9	49	140	426	1272	2395	3677	5018	4992	3680	2359	1191	405	115	35	5	0	0	0	1249.20
TYB 2N2N	6	12	58	110	268	516	1021	2027	3021	4018	5049	4966	3974	2988	2010	970	459	244	90	46	10	8	14.99
TBY 2NN	37	61	186	311	582	923	1491	2428	3305	4157	4963	5045	4207	3350	2445	1424	893	522	264	160	47	30	1331.37
TBY 3NN	86	127	312	474	768	1188	1767	2643	3475	4241	4980	5029	4287	3502	2656	1709	1120	748	431	269	111	70	4889.21
TBY 2N2N	5	8	46	107	235	484	998	1992	2977	3993	5021	4997	3991	3006	1977	948	483	243	90	45	6	3	14.77
TXA 2NN	14	21	64	115	239	451	871	1832	2851	3952	5167	4858	3605	2411	1277	358	83	13	3	0	0	0	669.11
TXA 3NN	2	8	29	57	152	298	664	1544	2605	3746	5060	4946	3610	2288	1066	225	40	6	0	0	0	0	1167.22
TXA 2N2N	36	50	143	214	429	743	1290	2308	3288	4227	5175	4839	3780	2754	1748	765	308	131	30	11	0	0	439.31
TAX 2NN	9	17	79	164	405	722	1305	2289	3185	4034	4813	5200	4396	3600	2728	1782	1184	790	485	336	141	95	3331.26
TAX 3NN	37	60	208	330	642	1008	1590	2526	3362	4117	4801	5207	4498	3790	2961	2021	1386	985	621	443	191	127	6862.90
TAX 2N2N	0	0	11	26	116	326	829	1843	2899	3943	4954	5056	4101	3151	2169	1199	682	391	211	132	42	21	292.49
TXB 2NN	15	27	74	126	261	473	914	1820	2901	3985	5184	4837	3613	2395	1257	349	87	16	1	0	0	0	684.91
TXB 3NN	5	7	36	64	141	311	670	1509	2637	3855	5078	4931	3568	2223	1052	244	47	5	1	0	0	0	1163.42
TXB 2N2N	38	66	166	251	456	765	1291	2339	3288	4250	5210	4806	3785	2747	1739	771	292	112	37	11	0	0	585.42
TBX 2NN	8	15	68	149	375	701	1241	2250	3150	3966	4767	5240	4414	3596	2748	1805	1222	851	538	368	179	128	4808.28
TBX 3NN	35	50	180	312	607	960	1541	2494	3302	4050	4788	5219	4516	3765	2941	2018	1417	1006	647	477	218	161	8405.30
TBX 2N2N	0	0	5	29	117	318	800	1790	2866	3931	4863	5154	4138	3166	2181	1225	734	441	245	169	52	35	501.19
TYA 2NN	1	1	10	22	97	237	617	1505	2595	3767	4991	5032	3781	2627	[illegible]	584	204	74	27	10	2	1	577.38
TYA 3NN	0	0	5	11	50	146	472	1309	2342	3605	4947	5067	3683	2372	1253	390	109	26	5	1	0	0	1206.50
TYA 2N2N	6	9	48	106	249	508	975	1997	3009	4067	5035	4982	3993	2989	2028	1026	504	237	90	56	14	6	22.09
TAY 2NN	50	78	218	335	628	981	1553	2476	3359	4172	5015	4998	4147	3316	2410	1413	840	494	245	141	33	21	1530.40
TAY 3NN	107	160	344	498	818	1219	1824	2709	3474	4236	5008	4996	4245	3491	2660	1701	1090	704	409	261	97	56	5508.67
TAY 2N2N	4	9	53	109	241	515	1021	2013	3047	4038	5059	4950	3953	2915	1924	959	456	217	71	29	7	1	23.62
TD 2N	8	18	82	130	287	564	1064	2011	3053	4032	5035	5003	3973	2916	1928	949	420	206	79	33	4	1	51.49
TX NN	3	6	49	89	238	513	1023	2062	3090	4057	5071	4951	4001	2983	2011	1016	490	240	91	42	5	1	19.79
TY NN	7	13	53	103	246	491	995	1990	2956	3995	4986	5046	3990	2983	1964	983	491	240	92	40	8	4	9.40

N= 128

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
TYY 2NN	3	7	55	106	236	475	956	1914	2940	3928	4954	5055	4045	3072	2048	1040	497	239	85	45	7	3	18.10
TYY 3NN	1	8	50	104	244	479	970	1958	2972	3962	4896	5110	4067	3087	2106	1019	495	230	90	44	5	0	29.34
TYY 2N2N	4	8	42	98	236	500	984	1921	2948	3949	4971	5049	4005	3046	2009	996	489	237	86	45	8	3	16.26
TXX 2NN	10	17	65	117	262	486	987	1905	2828	3826	4807	5202	4153	3122	2073	991	453	213	69	26	2	0	53.34
TXX 3NN	8	17	83	135	258	507	972	1903	2820	3782	4778	5229	4202	3101	2089	964	444	191	63	24	0	0	91.85
TXX 2N2N	6	13	46	100	253	502	1000	1935	2942	3943	4924	5086	4045	3049	2043	1056	516	238	75	31	7	4	24.57
TXY 2NN	15	28	102	169	333	590	1068	2022	2975	3993	4955	5058	4094	3055	1978	912	394	168	54	23	1	0	125.30
TXY 3NN	20	29	100	145	305	572	1056	1985	2983	3989	4944	5069	4124	3074	2008	912	405	189	58	26	7	1	130.68
TXY 2N2N	12	19	83	142	323	591	1083	2024	2983	3963	4987	5026	3977	2942	1971	936	423	18'	52	22	2	1	72.54
TYX 2NN	2	7	33	72	198	414	868	1870	2879	3837	4833	5181	4131	3165	2181	1151	593	293	115	59	10	8	56.25
TYX 3NN	2	6	44	84	226	429	876	1837	2871	3822	4808	5206	4172	3156	2128	1128	554	275	105	38	6	2	52.63
TYX 2N2N	1	2	23	62	206	413	873	1896	2876	3907	4920	5095	4076	3122	2139	1111	600	312	140	73	11	9	66.05
TYB 2NN	0	0	9	22	79	205	565	1424	2518	3732	4934	5085	3833	2646	1542	617	227	69	18	3	1	0	632.94
TYB 3NN	0	0	1	7	42	119	391	1171	2276	3600	4911	5097	3736	2471	1331	415	114	27	3	1	0	0	1267.00
TYB 2N2N	5	8	44	96	232	487	979	1948	2923	3939	4943	5071	4027	3050	2031	1001	490	254	98	50	13	5	11.20
TBY 2NN	32	67	171	295	537	852	1437	2440	3312	4112	4909	5096	4242	3439	2520	1517	899	524	282	173	55	31	1367.45
TBY 3NN	77	119	284	425	739	1119	1719	2638	3485	4184	4923	5081	4340	3623	2772	1772	1170	729	416	286	108	78	4632.66
TBY 2N2N	6	16	56	104	254	507	988	1936	2950	3982	4988	5022	4015	2999	2026	980	475	232	81	38	7	5	18.13
TXA 2NN	11	13	40	69	172	331	724	1611	2628	3731	4890	5125	3839	2584	1365	442	113	21	3	3	0	0	704.61
TXA 3NN	3	3	16	37	104	206	492	1305	2359	3533	4859	5149	3762	2333	1149	268	55	5	0	0	0	0	1381.48
TXA 2N2N	20	38	112	192	363	642	1154	2083	3064	3970	4995	5018	3974	2971	1936	883	382	159	40	14	2	2	199.57
TAX 2NN	15	28	108	174	382	689	1247	2209	3099	3965	4777	5231	4385	3596	2675	1730	1124	741	438	293	114	70	2329.20
TAX 3NN	48	73	205	317	572	955	1527	2460	3286	4015	4762	5246	4514	3741	2899	1984	1340	915	580	408	190	125	6306.78
TAX 2N2N	0	0	17	55	170	389	838	1863	2838	3820	4868	5142	4081	3108	2187	1189	668	358	185	92	27	16	159.66
TXB 2NN	10	18	44	82	193	372	765	1613	2664	3797	4987	5027	3794	2577	1403	432	117	26	2	0	0	0	645.44
TXB 3NN	2	2	25	40	98	206	526	1353	2375	3606	4957	5047	3676	2371	1123	303	55	8	2	0	0	0	1295.24
TXB 2N2N	26	41	126	213	393	674	1166	2109	3059	4030	5045	4966	3974	2955	1923	857	362	149	38	11	3	1	265.63
TBX 2NN	11	25	92	171	360	688	1232	2237	3132	3972	4784	5223	4378	3578	2714	1771	1180	801	475	336	134	91	3129.14
TBX 3NN	41	57	186	299	562	934	1477	2465	3313	4040	4800	5205	4458	3727	2923	1997	1396	984	620	447	202	149	7439.17
TBX 2N2N	0	1	17	49	159	362	830	1821	2829	3867	4893	5122	4141	3152	2207	1234	678	394	194	116	35	23	217.01
TYA 2NN	0	0	8	21	83	192	535	1428	2510	3680	4915	5103	3883	2718	1617	621	219	84	17	7	2	0	611.00
TYA 3NN	0	0	2	7	50	135	356	1167	2267	3513	4887	5116	3760	2502	1339	405	119	29	4	3	0	0	1286.06
TYA 2N2N	3	8	37	87	233	466	919	1958	2947	3891	4912	5099	4087	3104	2050	1033	532	259	105	52	9	6	21.69
TAY 2NN	55	80	196	301	555	909	1472	2444	3333	4120	4904	5106	4264	3412	2522	1501	863	488	255	146	45	26	1513.10
TAY 3NN	103	136	292	445	774	1132	1730	2656	3506	4237	4902	5107	4358	3597	2781	1765	1144	699	393	270	93	64	5008.59
TAY 2N2N	6	12	64	112	265	528	1026	1980	2952	3961	4963	5047	3977	2999	2004	975	470	226	77	34	7	2	23.58
TD 2N	16	25	76	123	295	553	1085	2042	3030	4011	5082	4950	3919	2960	1961	991	455	211	83	36	4	3	56.06
TX NN	2	5	43	95	222	484	982	1974	2972	3967	4933	5084	4131	3114	2062	1009	497	242	95	44	8	2	20.54
TY NN	6	10	51	92	232	488	954	1939	2955	3946	4917	5105	4044	3040	2053	1035	516	241	81	45	7	4	19.25

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
TYY 2NN	3	11	46	104	235	490	997	1995	3000	4013	5048	4960	3998	2959	1945	979	498	254	98	42	6	4	17.13
TYY 3NN	4	5	57	106	252	482	992	2023	2987	4037	5063	4943	3939	2956	1983	957	487	261	90	44	7	3	24.50
TYY 2N2N	8	14	50	103	261	479	975	1989	3024	4046	5051	4961	4032	3013	2034	1008	510	239	96	48	8	4	19.03
TXX 2NN	9	13	63	115	273	530	1051	1978	2954	3944	4916	5088	4050	3074	2065	955	447	199	81	38	5	2	40.22
TXX 3NN	13	23	79	126	288	537	1000	1974	2937	3912	4897	5111	4080	3067	1987	968	464	201	67	35	5	2	52.77
TXX 2N2N	4	5	35	88	267	519	1042	1973	2942	4012	5030	4983	4013	3017	2038	987	506	239	86	37	2	1	36.89
TXY 2NN	10	16	76	133	307	540	1063	2064	3086	4075	5046	4959	3952	2950	1976	941	448	202	75	34	5	3	36.69
TXY 3NN	15	25	79	135	284	530	1027	2050	3118	4009	5057	4948	3929	2952	1990	953	458	216	67	33	5	3	65.06
TXY 2N2N	5	8	67	123	281	550	1050	2073	3021	4015	4999	5013	3977	2983	1987	969	452	215	76	39	5	4	27.19
TYX 2NN	5	6	41	86	224	489	957	1927	2905	3908	4932	5078	4107	3059	2040	1020	514	263	94	50	7	6	21.04
TYX 3NN	6	11	48	111	244	498	968	1975	2931	3917	4950	5052	4088	3037	2011	998	516	255	98	43	8	3	19.00
TYX 2N2N	3	5	38	75	217	452	942	1935	2971	3971	5008	4999	3995	3014	2030	1036	544	263	131	64	9	6	30.39
TYB 2NN	0	0	7	22	71	217	596	1480	2562	3780	5046	4969	3729	2557	1485	574	211	75	20	5	1	0	648.34
TYB 3NN	0	0	2	4	41	136	412	1244	2343	3702	5026	4980	3587	2318	1219	383	119	29	7	2	0	0	1314.48
TYB 2N2N	6	10	47	90	242	505	987	1970	3023	4010	5033	4974	3989	2996	1987	1016	512	231	95	45	7	4	14.20
TBY 2NN	37	67	173	272	530	883	1503	2476	3375	4220	5035	4974	4145	3329	2431	1473	861	541	274	183	52	29	1340.00
TBY 3NN	79	115	301	443	734	1150	1764	2683	3532	4310	5072	4935	4207	3466	2666	1720	1117	757	431	285	113	71	4622.25
TBY 2N2N	8	16	59	96	254	503	968	2007	3010	4024	5008	4997	4014	3022	2052	1016	490	226	92	53	10	4	21.68
TXA 2NN	1	5	19	45	143	319	717	1609	2662	3826	5007	5006	3752	2548	1447	472	133	38	5	4	0	0	607.70
TXA 3NN	1	2	11	27	82	198	489	1342	2381	3625	4975	5027	3651	2297	1135	303	84	16	3	1	0	0	1307.42
TXA 2N2N	9	17	65	150	343	628	1166	2077	3051	4047	5062	4944	3942	2941	1958	912	443	184	45	13	2	1	99.34
TAX 2NN	21	38	118	221	451	814	1397	2321	3232	4072	4897	5107	4314	3504	2618	1583	1020	634	360	241	93	62	1804.58
TAX 3NN	66	91	236	373	661	1053	1637	2584	3406	4142	4885	5117	4408	3650	2841	1815	1240	837	515	352	138	96	5069.06
TAX 2N2N	2	3	26	56	189	448	924	1939	2935	3918	4911	5094	4125	3095	2075	1105	584	312	142	79	15	10	55.83
TXB 2NN	3	5	19	39	144	316	705	1654	2707	3868	5070	4936	3737	2516	1441	483	141	45	7	4	0	0	580.55
TXB 3NN	2	4	13	23	71	188	482	1335	2444	3749	5018	4987	3598	2310	1164	336	85	24	4	1	0	0	1237.13
TXB 2N2N	10	17	84	151	322	634	1140	2120	3082	4025	5021	4984	3964	2907	1923	940	431	170	56	21	2	1	88.64
TBX 2NN	16	26	111	213	449	784	1369	2314	3256	4077	4873	5134	4319	3470	2636	1584	1017	666	375	253	97	59	1820.04
TBX 3NN	54	87	226	357	661	1051	1635	2582	3416	4160	4858	5144	4393	3603	2784	1830	1237	855	546	364	155	116	5606.49
TBX 2N2N	0	1	32	60	184	422	894	1903	2917	3951	4917	5090	4077	3063	2100	1111	583	315	152	80	25	12	73.07
TYA 2NN	0	0	6	17	75	209	583	1530	2577	3733	4963	5054	3788	2582	1494	582	203	74	13	7	0	0	630.57
TYA 3NN	0	0	4	6	41	139	413	1233	2354	3647	4994	5011	3641	2350	1182	377	122	29	3	1	0	0	1324.03
TYA 2N2N	5	8	41	79	234	496	967	1974	3017	4000	5023	4983	3955	2969	1967	1024	519	251	111	52	10	8	18.12
TAY 2NN	43	74	192	285	553	920	1501	2489	3375	4243	5071	4932	4126	3310	2439	1475	895	537	279	169	50	27	1451.31
TAY 3NN	88	125	313	442	753	1150	1766	2725	3525	4282	5059	4943	4197	3440	2661	1747	1140	751	429	276	111	68	4836.81
TAY 2N2N	9	14	60	110	266	492	1013	2005	3045	4024	5007	5003	4040	3049	2014	1017	486	243	91	48	13	8	18.74
TD 2N	2	7	69	110	290	533	1014	1993	3015	3985	5008	5011	4013	3003	1965	978	512	231	91	44	9	3	34.92
TX NN	2	7	52	104	256	504	1018	1990	2967	3956	4943	5070	4026	2972	1349	980	452	204	83	42	9	4	20.80
TY NN	5	7	52	107	248	501	1002	2041	3019	4053	5032	4984	3990	3008	1969	931	483	237	90	50	7	4	19.75

N= 512

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
TYY 2NN	5	9	44	93	247	497	975	1921	2938	3921	4937	5074	4037	2998	1971	983	470	237	90	40	8	5	12.86
TYY 3NN	3	8	42	84	241	475	961	1924	2870	3904	4955	5049	4029	3012	1961	990	494	235	89	39	5	3	20.58
TYY 2N2N	4	10	53	121	273	508	1011	1984	2975	3962	4994	5012	3996	3012	1963	984	473	252	113	47	9	5	23.77
TXX 2NN	5	10	67	117	294	550	1028	2024	2953	3946	4942	5062	4042	3059	2018	1020	513	266	108	43	9	4	26.80
TXX 3NN	8	19	73	122	301	566	1035	1993	2953	3928	4929	5074	4101	3023	2075	1053	511	272	100	42	6	2	47.34
TXX 2N2N	1	10	49	101	254	507	1010	2043	3001	3989	4980	5026	3962	2942	1966	1000	498	243	105	61	16	7	22.03
TXY 2NN	10	16	60	112	276	502	1018	1974	2964	3987	4953	5050	3985	2928	2008	953	466	202	74	32	6	2	35.86
TXY 3NN	6	13	63	108	263	515	1025	1963	2948	3972	4934	5072	3982	2943	1963	966	454	215	84	36	3	1	29.69
TXY 2N2N	8	16	71	117	274	512	1040	2005	3013	4032	5050	4956	3961	2922	1906	912	414	212	86	46	13	6	33.06
TYX 2NN	6	11	52	102	256	499	1002	1957	2938	3908	4933	5072	4079	3081	2137	1074	540	294	129	63	15	6	25.53
TYX 3NN	7	9	48	90	253	515	965	1943	2916	3929	4890	5114	4130	3093	2082	1048	540	295	122	58	9	4	26.54
TYX 2N2N	2	6	46	103	232	493	972	1987	2987	3996	4956	5048	4065	3058	2010	1051	549	280	120	66	13	7	21.41
TYB 2NN	0	2	9	21	89	219	566	1462	2586	3756	4927	5084	3839	2618	1507	545	203	73	14	4	1	1	651.76
TYB 3NN	0	0	4	7	34	114	388	1214	2329	3531	4920	5084	3724	2354	1272	382	111	27	4	2	0	0	1337.97
TYB 2N2N	3	5	57	125	283	481	1023	1970	2983	3970	4979	5025	4045	3016	1990	962	493	255	103	48	7	3	37.48
TBY 2NN	31	47	165	265	531	881	1431	2395	3252	4109	4946	5058	4210	3344	2425	1434	890	523	291	172	59	30	1186.04
TBY 3NN	62	99	276	430	715	1097	1677	2603	3425	4211	4947	5057	4291	3524	2693	1697	1116	763	446	282	117	71	4135.84
TBY 2N2N	10	15	56	109	258	523	996	1996	2986	3952	4973	5031	3993	2957	1984	939	488	246	103	56	14	9	23.40
TXA 2NN	0	1	21	45	125	303	680	1622	2648	3776	5013	4993	3766	2536	1430	518	167	46	9	4	0	0	591.61
TXA 3NN	0	0	8	19	72	177	483	1328	2399	3635	4953	5054	3695	2376	1210	358	97	16	1	0	0	0	1197.40
TXA 2N2N	10	19	87	136	302	589	1093	2088	3049	4056	5074	4927	3927	2911	1918	910	416	183	75	42	4	2	59.77
TAX 2NN	29	51	152	266	503	825	1424	2350	3275	4143	4923	5076	4276	3488	2595	1615	1019	675	381	264	103	74	2327.30
TAX 3NN	67	100	281	422	724	1073	1650	2597	3426	4212	4940	5061	4355	3628	2828	1898	1276	885	530	363	184	127	6815.46
TAX 2N2N	1	4	35	81	203	442	941	1955	2970	3976	4962	5043	4030	3010	2047	1038	581	322	141	71	27	16	64.98
TXB 2NN	0	2	22	44	114	268	649	1575	2621	3837	5035	4969	3753	2516	1384	474	149	41	11	5	0	0	677.74
TXB 3NN	0	0	6	12	57	157	448	1285	2375	3628	4973	5029	3667	2311	1149	343	83	13	2	0	0	0	1336.12
TXB 2N2N	8	24	87	147	321	589	1084	2037	3068	4062	5069	4934	3938	2887	1876	881	392	176	70	33	7	2	80.28
TBX 2NN	21	40	139	250	500	830	1398	2360	3271	4085	4912	5092	4276	3472	2617	1653	1030	683	389	271	116	81	2487.72
TBX 3NN	58	97	256	413	691	1061	1620	2593	3406	4186	4913	5088	4351	3612	2829	1897	1271	884	535	383	180	127	6578.70
TBX 2N2N	1	3	39	85	194	438	946	1939	2963	3973	4956	5050	4109	3059	2054	1085	591	340	137	71	27	17	83.78
TYA 2NN	0	1	13	31	88	245	606	1524	2566	3747	4922	5089	3825	2670	1562	600	235	87	21	7	0	0	541.36
TYA 3NN	0	0	6	9	38	126	414	1250	2296	3549	4943	5060	3710	2444	1291	428	126	41	4	1	0	0	1221.65
TYA 2N2N	4	6	60	117	255	508	1002	2009	2997	4003	4994	5012	4034	3029	2023	1043	512	249	125	61	7	3	27.33
TAY 2NN	33	53	177	276	523	880	1450	2403	3289	4156	4974	5030	4203	3351	2400	1421	885	549	273	157	56	31	1201.24
TAY 3NN	83	115	275	425	748	1112	1685	2610	3460	4249	4973	5031	4302	3530	2657	1686	1146	751	428	272	108	72	4464.13
TAY 2N2N	10	12	62	117	256	510	1019	1993	2976	4021	5020	4983	3964	2941	1931	939	460	225	94	58	19	7	33.56
TD 2N	8	13	69	144	309	551	1054	2056	3044	3996	5000	5014	4031	3056	2019	1037	483	238	83	39	8	3	37.47
TX NN	3	8	43	95	254	495	1044	2035	3047	4013	5034	4976	3971	2981	2027	1017	522	256	108	54	9	6	14.17
TY NN	5	9	49	98	221	492	1004	1995	3034	3960	4965	5045	4035	3010	1981	974	471	223	[illegible]	44	4	1	19.05

N=1024

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
TYY 2NN	6	11	50	98	249	492	962	1975	3001	4037	5006	4997	4002	3016	2033	1002	513	224	102	55	12	7	19.75
TYY 3NN	6	13	44	90	225	486	992	1996	2971	4007	5004	4998	4005	3029	2053	1010	503	255	97	46	11	7	12.60
TYY 2N2N	4	10	45	103	237	494	976	1953	2950	4016	5024	4983	3970	2962	1957	1008	521	275	98	49	8	3	18.66
TXX 2NN	5	13	60	103	257	504	995	1963	2942	3916	4922	5083	4089	3034	2010	993	479	245	89	49	15	6	18.38
TXX 3NN	6	9	58	113	266	505	999	1967	2896	3917	4924	5079	4053	3007	2037	999	488	221	91	47	9	4	20.82
TXX 2N2N	3	5	54	101	257	484	997	2009	2994	3984	4953	5054	4038	2982	1991	1046	541	277	107	53	7	1	23.68
TXY 2NN	4	14	66	114	279	529	992	2004	3042	3997	5012	4995	3999	3004	1989	965	462	215	89	50	10	7	22.46
TXY 3NN	8	12	56	113	287	531	1023	2037	2984	4010	4998	5004	4006	2987	2008	996	478	215	91	41	12	6	21.55
TXY 2N2N	5	9	53	117	280	526	1043	2018	3032	4042	5011	4993	3938	2941	1981	966	493	257	97	45	6	1	19.16
TYX 2NN	5	8	38	86	217	480	943	1875	2929	3912	4905	5097	4073	3043	2024	1026	514	290	113	62	16	7	31.66
TYX 3NN	5	9	45	98	231	479	949	1909	2863	3924	4925	5083	4073	3047	2039	1018	516	274	112	56	11	8	18.35
TYX 2N2N	4	10	38	74	228	456	936	1925	2910	3917	4930	5073	3998	2980	1984	1012	558	296	123	64	9	3	34.40
TYB 2NN	1	2	12	27	81	213	563	1446	2555	3796	5039	4965	3766	2618	1503	599	217	76	24	11	0	0	638.23
TYB 3NN	0	2	5	8	31	118	363	1252	2344	3643	4989	5013	3620	2399	1281	400	129	30	8	2	0	0	1278.99
TYB 2N2N	7	14	52	108	252	476	961	1955	2987	3976	5020	4984	3962	2950	2001	1016	514	248	91	53	10	2	20.19
TBY 2NN	30	53	164	279	514	869	1481	2478	3349	4158	5006	4998	4181	3367	2490	1484	904	556	280	170	60	42	1337.67
TBY 3NN	75	107	269	430	747	1149	1732	2679	3512	4240	5002	4998	4257	3537	2736	1745	1139	767	450	292	117	74	4580.77
TBY 2N2N	4	8	40	88	236	494	979	1963	2970	3976	5060	4946	3939	2930	1981	999	516	268	113	58	5	2	20.82
TXA 2NN	0	1	18	38	117	268	634	1585	2630	3802	4979	5027	3828	2396	1457	533	188	64	18	7	1	0	580.95
TXA 3NN	0	0	5	10	59	160	452	1280	2343	3603	4970	5032	3688	2380	1208	337	103	27	5	0	0	0	1278.73
TXA 2N2N	7	15	73	140	308	554	1062	2049	3002	4023	4971	5030	3999	2943	1966	973	490	249	84	33	5	2	37.67
TAX 2NN	25	39	135	250	493	821	1387	2332	3218	4049	4909	5099	4272	3440	2528	1552	953	595	326	212	81	57	1545.61
TAX 3NN	64	98	260	403	694	1071	1632	2550	3368	4142	4901	5102	4335	3574	2737	1803	1204	799	485	314	143	101	5040.60
TAX 2N2N	1	2	27	76	205	463	909	1939	2932	3926	4946	5055	4054	3048	2056	1076	575	310	138	58	12	4	47.97
TXB 2NN	2	4	19	40	113	282	632	1552	2648	3825	5006	5002	3795	2587	1456	522	195	58	17	10	0	0	590.70
TXB 3NN	1	1	6	15	53	160	443	1282	2409	3669	4978	5028	3658	2348	1214	347	98	33	7	3	0	0	1253.72
TXB 2N2N	7	18	81	137	309	572	1081	2056	3018	4064	5043	4958	3945	2958	1975	941	466	236	87	42	5	1	39.41
TBX 2NN	17	38	131	235	470	824	1356	2313	3234	4093	4897	5105	4279	3422	2525	1549	958	601	350	241	87	58	1632.95
TBX 3NN	61	88	244	380	677	1053	1623	2520	3391	4163	4897	5104	4363	3565	2743	1812	1196	812	477	342	161	123	5691.56
TBX 2N2N	2	5	26	61	219	429	896	1893	2899	3882	4942	5062	4048	3006	2026	1055	562	329	141	75	16	8	60.60
TYA 2NN	0	1	10	20	76	215	553	1469	2543	3748	4962	5045	3784	2622	1512	597	221	82	26	8	1	0	630.49
TYA 3NN	0	0	6	8	36	121	383	1175	2302	3592	4969	5032	3693	2379	1272	391	118	34	10	4	0	0	1330.67
TYA 2N2N	9	12	46	98	232	470	939	1970	2960	3930	4999	5005	4008	2991	1991	1001	518	270	104	55	9	4	19.47
TAY 2NN	32	66	175	284	535	926	1483	2446	3353	4189	4988	5012	4166	3350	2504	1481	882	540	274	158	55	35	1370.32
TAY 3NN	74	114	292	427	774	1172	1744	2686	3505	4227	5009	4992	4243	3533	2718	1750	1134	756	428	283	110	70	4534.66
TAY 2N2N	3	6	38	89	258	521	969	1998	2971	3980	5068	4935	3923	2948	1979	985	490	264	108	48	7	0	33.00
TD 2N	4	13	58	107	275	521	979	2004	3021	4016	5028	4988	3908	2911	1966	943	449	204	84	40	11	5	29.62
TX NN	3	7	47	93	238	480	995	1957	2934	3934	4947	5063	4022	3017	2031	1032	508	238	109	58	11	3	15.28
TY NN	4	9	55	98	244	477	989	1986	2994	3979	4962	5046	4032	3027	2028	1011	522	251	116	58	15	5	14.80

N= 2

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
FYY	NNON	3	8	52	95	230	456	997	1996	3047	4057	5078	4922	3953	2931	1957	1001	491	244	103	50	9	7	20.13
FYY	NNNN	25	40	104	158	293	508	971	2039	3125	4128	5115	4885	3944	2957	1972	980	489	272	134	84	38	27	250.37
FXX	NNON	85	85	87	90	148	288	565	1114	1730	2881	4540	5460	2953	1852	1332	736	442	295	194	139	71	57	5772.78
FXX	NNNN	20	20	39	70	150	247	431	961	1602	2848	5078	4922	2575	1648	1137	667	430	322	243	194	117	93	6565.06
FXY	NNON	10	16	78	171	350	609	1024	1680	2378	3355	4901	5099	3878	2995	2363	1648	1162	804	471	305	30	47	2747.48
FXY	NNNN	18	39	165	248	461	697	991	1559	2329	340[illegible]	5043	4957	3498	2658	2136	1550	1166	873	578	414	153	92	5485.37
FYX	NNON	5	10	51	92	203	413	833	1790	2794	3867	4911	5069	4035	3141	2244	1325	751	460	231	130	28	16	312.84
FYX	NNNN	4	6	48	112	254	486	900	1853	2952	4061	5060	4940	3930	2982	2111	1202	693	417	231	159	60	38	457.96
FYB	NNON	5	12	51	106	239	518	1065	2099	3102	4071	5011	4989	4048	3124	2131	1107	601	310	127	66	12	6	49.08
FYB	NNNN	30	53	146	220	348	577	1046	2102	3173	4171	5067	4933	4020	3018	2099	1104	595	336	162	117	42	30	481.86
FBY	NNON	5	13	60	112	262	518	1059	2140	3130	4033	5005	4995	4093	3209	2238	1195	653	342	152	76	17	10	106.64
FBY	NNNN	17	31	103	172	338	608	1168	2218	3184	4114	5034	4966	4102	3180	2251	1238	675	373	173	112	43	27	312.94
FXA	NNON	23	24	24	29	61	104	218	464	892	1752	3379	6621	4506	3138	2297	1388	852	518	294	179	78	57	4268.20
FXA	NNNN	2	4	13	21	36	82	161	439	932	2069	4194	5806	3765	2525	1825	1105	696	450	301	233	124	89	5811.10
FAX	NNON	24	25	29	39	77	150	312	643	1064	1932	3373	6627	4670	3269	2483	1599	1050	672	376	230	96	68	4225.10
FAX	NNNN	4	5	14	30	67	134	260	579	1070	2087	3782	6218	4153	2798	2146	1467	997	701	425	282	142	106	6011.41
FXB	NNON	13	13	40	68	185	329	627	1185	1802	2720	4539	5461	3953	3200	2535	1718	1127	733	371	221	53	27	2843.57
FXB	NNNN	13	16	53	102	212	347	593	1054	1702	2850	4790	5210	3450	2716	2222	1550	1102	752	459	309	107	74	4685.18
FBX	NNON	1	4	21	48	118	252	536	1200	2053	3238	4538	5462	4386	3470	2562	1501	863	489	231	119	25	15	853.49
FBX	NNNN	2	5	27	40	121	257	557	1246	2262	3566	4806	5194	4171	3280	2366	1384	794	466	249	167	69	49	1076.17
FYA	NNON	2	11	49	98	247	531	1024	2077	3048	4033	4984	5016	4086	3181	2267	1262	750	464	255	146	33	20	326.79
FYA	NNNN	14	28	109	173	351	603	1066	2067	3157	4158	5030	4970	4011	3032	2141	1173	677	406	228	154	59	40	504.21
FAY	NNON	17	28	155	273	503	816	1304	2086	2885	3818	4897	5103	4153	3339	2545	1730	1265	919	588	419	156	95	4620.31
FAY	NNNN	63	106	281	407	667	947	1368	2110	2944	3913	5011	4989	3992	3137	2413	1705	1254	990	701	544	272	189	11613.63
RXY	2N2N	608	1086	2932	3917	5053	5622	6019	6305	6448	6602	6814	3191	2909	2501	1942	1177	657	353	148	76	19	12	227166.84
RX	2N2N	720	1105	2021	2245	2419	2544	2787	3296	3819	4389	5122	5095	4381	3810	3331	2816	2586	2465	2261	2057	1155	748	314767.02
RY	2N2N	4	9	58	109	246	482	1034	2029	2994	3994	4964	5052	4036	3086	2127	1085	523	254	93	52	17	5	40.25
FX	NONN	78	78	81	85	137	289	554	1118	1704	2873	4514	5486	2981	1821	1303	716	425	276	180	131	68	50	5454.34
FY	NONN	7	12	55	110	251	490	1013	2020	3022	4017	5013	4987	3985	3005	1975	973	457	232	88	46	12	11	20.84
FX	2NNN	46	46	51	57	120	249	463	1060	2376	4013	5058	4942	4134	3087	1367	543	331	214	158	128	94	85	3529.32
FY	2NNN	4	12	60	112	279	544	1078	2068	3061	4086	5140	4860	3871	2903	2025	980	482	228	105	65	12	8	43.72

N= 4

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
FYY	NNON	3	3	42	90	238	481	979	1957	2928	3953	4948	5052	3999	3015	2009	1010	497	243	105	52	15	7	15.58
FYY	NNNN	3	7	53	97	243	480	999	1933	2981	3982	5012	4988	4023	3039	2066	1047	522	272	117	53	14	7	21.22
FXX	NNON	20	21	38	59	147	268	519	854	1397	2542	4025	5975	4064	2207	1314	520	187	97	39	16	4	3	3070.87
FXX	NNNN	2	4	31	55	106	172	285	700	1476	2714	4583	5417	3484	2123	1242	517	178	93	48	28	4	4	3100.89
FXY	NNON	11	14	59	128	296	517	887	1586	2398	3421	4803	5197	3896	2939	2050	1149	746	510	326	229	113	87	2091.03
FXY	NNNN	7	19	101	160	313	482	732	1393	2420	3643	5009	4991	3666	2672	1870	1032	609	447	310	230	121	94	2540.05
FYX	NNON	1	1	32	80	211	415	842	1768	2748	3748	4788	5212	4172	3131	2111	1147	622	378	183	122	40	24	202.10
FYX	NNNN	3	5	46	84	213	424	860	1775	2862	3882	4954	5046	4063	3069	2122	1106	614	345	175	95	29	16	122.63
FYB	NNON	2	8	37	81	249	527	1044	2061	3012	4008	4982	5018	4021	3053	2076	1088	588	319	152	81	25	11	58.70
FYB	NNNN	6	10	58	118	275	540	1071	2054	3065	4066	5053	4947	3995	3092	2156	1109	626	348	151	82	22	10	76.21
FBY	NNON	10	16	66	126	269	560	1100	2177	3148	4165	5125	4875	3931	3067	2136	1206	670	358	166	93	29	20	162.81
FBY	NNNN	10	17	66	129	311	592	1179	2263	3293	4248	5186	4814	3946	3063	2190	1245	730	419	204	112	32	24	293.89
FXA	NNON	3	5	10	20	44	105	219	412	916	1694	3026	6974	5320	4230	3092	1650	947	544	260	157	44	26	2926.10
FXA	NNNN	0	1	12	20	46	76	129	430	1139	2269	3853	6147	4465	3355	2374	1245	625	332	160	96	16	12	2279.04
FAX	NNON	4	4	14	23	64	133	253	512	1128	2106	3132	6868	5873	4651	3164	1917	1219	837	540	378	123	76	4665.22
FAX	NNNN	0	0	12	30	58	104	194	674	1502	2517	3499	6501	5380	4019	2663	1576	971	648	407	296	122	78	3223.66
FXB	NNON	2	6	23	54	146	255	509	1016	1600	2458	4003	5997	4376	3339	2586	1643	1151	808	512	360	155	96	4734.93
FXB	NNNN	1	4	19	36	93	191	365	798	1671	2901	4424	5576	3923	2860	2137	1308	856	622	425	325	151	111	4615.89
FBX	NNON	3	5	17	39	106	215	480	1184	2104	3100	4114	5886	4884	3947	2910	1754	1040	627	316	175	59	34	1326.58
FBX	NNNN	3	4	16	33	94	228	539	1333	2335	3393	4379	5621	4680	3732	2692	1566	916	525	251	143	37	22	774.12
FYA	NNON	4	9	63	106	258	529	1062	2026	3045	4017	4993	5007	4034	3078	2122	1128	633	361	180	110	36	25	150.20
FYA	NNNN	4	12	60	111	289	562	1054	2067	3091	4069	5064	4936	4016	3082	2133	1159	636	359	169	93	27	17	105.24
FAY	NNON	11	24	110	205	433	711	1213	2151	3142	4169	5169	4831	3871	2995	2152	1263	774	510	305	214	107	84	1695.74
FAY	NNNN	29	57	184	280	475	746	1200	2260	3326	4352	5352	4648	3743	2921	2138	1270	832	560	333	238	119	92	2442.17
RXY	2N2N	3563	3802	4135	4216	4297	4371	4495	4769	5048	5422	5859	4142	3669	3110	2407	1511	901	550	262	141	39	23	2548233.72
RX	2N2N	2358	2417	2524	2574	2686	2828	3110	3601	4075	4558	5008	5012	4555	4072	3599	3078	2841	2685	2576	2529	2429	2369	2228945.63
RY	2N2N	3	7	44	91	248	515	1042	2051	2991	3993	4988	5015	4065	3053	2082	1067	538	260	83	42	6	2	30.43
FX	NONN	22	22	32	52	156	288	489	838	1330	2459	3960	6040	4130	2204	1316	517	202	91	35	12	4	4	3289.95
FY	NONN	1	10	53	94	233	495	950	1943	2878	3869	4893	5107	4117	3049	2047	1055	506	253	98	45	14	9	33.81
FX	2NNN	8	8	16	42	115	218	398	918	1877	3349	4775	5225	4080	2823	1489	595	352	138	37	21	6	3	1414.52
FY	2NNN	9	11	52	90	243	478	567	1936	2950	3949	4973	5027	4043	3008	2008	1012	525	268	106	49	6	5	18.24

N= 8

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
FYY	NNON	6	7	50	110	269	514	1033	2030	3079	4049	5021	4979	3995	3019	1980	962	477	271	111	44	7	3	29.62
FYY	NNNN	3	6	52	97	275	535	1050	2028	3033	3997	4999	5001	3995	3005	1989	1021	541	277	126	59	14	8	22.30
FXX	NNON	6	8	37	74	184	338	620	1144	1829	2866	4194	5806	4336	2731	1590	595	234	85	37	15	3	2	1370.27
FXX	NNNN	3	5	30	60	150	262	459	1086	2064	3309	4713	5287	3899	2683	1576	734	306	156	55	30	5	4	984.39
FXY	NNON	3	9	42	90	257	499	899	1693	2627	3648	4816	5184	4063	2999	1912	987	553	355	222	160	82	57	813.11
FXY	NNNN	6	11	52	109	227	402	706	1537	2613	3787	4943	5057	3920	2870	1900	973	531	305	180	128	65	50	675.31
FYX	NNON	6	12	42	87	230	464	964	1935	2888	3907	4916	5084	4094	3055	2055	1075	565	323	151	91	32	19	76.90
FYX	NNNN	3	6	50	94	244	495	1001	1937	2883	3951	4951	5049	4019	2980	2062	1069	565	295	142	82	22	10	52.12
FYB	NNON	4	11	36	90	264	561	1083	2146	3168	4142	5087	4913	3889	2922	2011	1036	568	307	147	93	20	14	86.10
FYB	NNNN	3	7	60	121	289	572	1129	2177	3186	4140	5124	4876	3908	3002	2023	1057	603	344	170	99	20	14	108.51
FBY	NNON	10	15	54	118	290	590	1173	2255	3288	4256	5189	4811	3856	2910	2032	1124	653	383	181	90	24	15	172.59
FBY	NNNN	6	14	75	143	341	669	1270	2359	3354	4317	5261	4739	3849	2933	2050	1188	713	438	216	127	36	22	351.48
FXA	NNON	1	6	41	67	166	317	646	1315	1981	2774	3953	6047	4796	3746	2717	1702	1052	633	317	184	65	44	1485.10
FXA	NNNN	6	8	34	64	157	299	562	1194	2141	3276	4470	5530	4327	3223	2251	1194	714	419	195	107	25	20	609.09
FAX	NNON	2	6	48	81	196	375	741	1566	2559	3394	4066	5934	5311	4527	3541	2084	1188	736	435	348	196	151	6188.47
FAX	NNNN	3	7	50	88	184	343	757	1850	2773	3544	4224	5776	5026	4150	3023	1698	926	552	294	202	108	82	2109.92
FXB	NNON	4	6	50	96	237	448	838	1492	2179	3109	4258	5742	4523	3461	2304	1339	904	671	476	355	207	158	5840.53
FXB	NNNN	8	13	51	95	222	347	647	1277	2235	3415	4598	5402	4215	3099	2116	1069	677	433	283	225	119	101	2414.77
FBX	NNON	1	6	42	103	271	510	968	1906	2809	3671	4483	5517	4702	3869	2963	1929	1225	781	411	268	80	54	1997.36
FBX	NNNN	10	15	66	127	286	546	1084	2038	2917	3768	4557	5443	4616	3759	2839	1745	1049	601	321	202	60	34	1112.76
FYA	NNON	8	14	55	107	284	558	1101	2136	3143	4147	5131	4869	3893	2942	2010	1078	598	309	150	97	27	16	88.24
FYA	NNNN	4	7	53	105	294	581	1134	2188	3192	4124	5139	4861	3882	2976	2048	1101	617	342	171	95	23	13	111.14
FAY	NNON	12	23	85	158	356	634	1155	2234	3302	4311	5297	4703	3774	2862	2027	1132	678	436	227	147	61	42	532.28
FAY	NNNN	14	26	102	171	331	573	1174	2359	3456	4465	5429	4571	3672	2835	2053	1184	737	462	259	165	81	56	974.27
RXY	2N2N	1875	1917	2093	2203	2443	2714	3154	3655	4461	4998	5554	4447	3871	3297	2550	1710	1173	795	443	311	121	79	703766.04
RX	2N2N	1666	1720	1939	2092	2374	2687	3061	3661	4139	4571	4981	5020	4630	4195	3739	3161	2728	2405	2131	1979	1733	1671	1112873.92
RY	2N2N	1	3	37	98	244	515	1038	2052	3029	3966	4970	5032	4049	3055	2054	1019	524	259	119	52	11	5	25.16
FX	NONN	5	4	36	69	179	347	623	1113	1804	2668	4183	5617	4355	2795	1644	605	257	90	32	16	4	3	1334.04
FY	NONN	5	12	54	106	256	501	1009	1992	2990	3948	4977	5023	4027	3013	1972	999	518	266	104	51	13	6	9.95
FX	2NNN	3	5	33	68	162	324	616	1264	2114	3266	4607	5393	4037	2817	1674	765	384	185	89	45	11	6	698.75
FY	2NNN	2	7	54	103	253	519	1064	2055	3022	4032	5002	4998	4004	3009	2032	1031	519	253	98	54	9	5	13.90

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
FYY NNON	3	10	56	107	244	495	1027	2016	2982	3979	4976	5024	4020	3016	2025	979	480	221	92	46	8	5	13.71
FYY NNNN	7	11	48	101	247	501	1025	2010	3028	4022	4987	5013	3987	2991	1978	976	467	240	90	46	8	1	12.50
FXX NNON	5	8	44	99	242	439	830	1690	2564	3514	4492	5508	4413	3218	1950	821	343	145	53	28	3	2	268.34
FXX NNNN	5	12	57	90	206	408	817	1673	2615	3601	4660	5340	4171	3025	1834	865	407	175	67	41	13	9	190.89
FXY NNON	5	14	42	92	243	466	920	1851	2797	3781	4795	5205	4166	3124	2036	1005	530	292	151	103	49	38	266.42
FXY NNNN	2	5	46	89	217	446	882	1821	2793	3829	4872	5128	4063	2973	1969	982	493	247	105	67	29	23	105.66
FYX NNON	0	3	34	81	239	463	946	1866	2821	3827	4830	5170	4106	3150	2123	1044	546	270	119	70	16	6	46.02
FYX NNNN	2	3	45	83	216	433	928	1928	2955	3952	4967	5033	4074	3127	2046	1022	494	259	113	61	12	7	32.38
FYB NNON	3	12	56	116	269	517	1026	2077	3118	4132	5084	4916	3905	2977	2008	1048	538	311	133	61	15	13	54.01
FYB NNNN	4	8	49	97	245	552	1096	2157	3201	4182	5143	4857	3903	2962	2006	1064	579	290	120	73	15	9	56.72
FBY NNON	4	10	51	110	284	575	1143	2246	3270	4271	5223	4777	3808	2897	2022	1054	579	334	152	84	30	16	120.83
FBY NNNN	4	10	65	136	342	677	1244	2330	3389	4367	5330	4670	3750	2866	2017	1096	637	378	176	106	37	28	295.32
FXA NNON	4	9	52	112	244	487	964	1901	2805	3716	4570	5430	4566	3615	2439	1247	660	446	313	219	85	52	1017.47
FXA NNNN	8	14	54	98	242	488	956	1866	2782	3696	4608	5392	4307	3218	2128	1047	543	269	136	90	39	23	161.78
FAX NNON	6	13	49	96	262	519	1027	2102	3114	4070	5022	4978	4070	3329	2675	2008	1465	950	451	280	95	69	3142.22
FAX NNNN	6	16	70	117	250	531	1101	2170	3241	4215	5142	4858	4035	3275	2602	1857	1238	762	355	207	56	32	1693.60
FXB NNON	6	9	56	103	252	462	960	1883	2813	3758	4660	5340	4444	3505	2467	1267	694	437	269	201	126	102	2092.38
FXB NNNN	6	6	45	90	249	495	935	1884	2808	3698	4681	5319	4259	3221	2130	1053	561	306	152	90	44	37	263.26
FBX NNON	5	11	46	96	250	543	1047	2076	3035	4008	4920	5080	4152	3328	2575	1749	1247	909	560	357	131	67	3239.87
FBX NNNN	7	11	55	101	263	562	1135	2231	3210	4139	5083	4917	4088	3276	2510	1689	1174	793	422	251	92	52	1920.50
FYA NNON	5	8	54	103	270	520	1033	2064	3071	4085	5037	4963	3966	2994	2029	1040	567	300	139	64	17	12	36.58
FYA NNNN	6	11	53	103	275	535	1091	2122	3169	4177	5110	4890	3904	2976	2040	1069	576	304	120	70	16	12	53.68
FAY NNON	0	6	46	109	281	551	1128	2224	3302	4291	5276	4724	3763	2847	1969	1104	619	365	187	122	48	35	314.64
FAY NNNN	3	9	50	109	299	571	1186	2328	3403	4468	5397	4603	3669	2803	1993	1142	668	402	231	136	52	38	485.58
RXY 2N2N	1082	1200	1571	1797	2165	2544	3096	3839	4469	5041	5549	4452	3878	3308	2664	1838	1327	953	624	443	209	159	246147.11
RX 2N2N	1184	1307	1713	1934	2290	2652	3116	3725	4231	4667	5044	4956	4554	4103	3627	3045	2632	2264	1899	1661	1279	1147	555342.40
RY 2N2N	6	6	41	98	244	500	1034	2047	3069	4080	5028	4973	3966	2975	2008	1059	547	286	110	65	9	5	25.89
FX NONN	4	10	40	83	230	420	813	1653	2581	3551	4525	5475	4300	3135	1930	796	355	164	57	26	3	2	244.09
FY NONN	3	10	52	108	248	497	1007	2051	3062	4055	5109	4891	3969	2973	1965	993	490	246	91	37	9	7	22.11
FX 2NNN	3	7	45	102	226	457	873	1776	2674	3674	4716	5284	4224	3056	1890	859	429	234	78	46	13	11	142.91
FY 2NNN	6	14	45	97	239	512	1027	2033	3037	4002	5025	4975	3951	2932	1924	944	443	210	87	37	6	3	19.95

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
FYY NNON	5	9	59	112	258	487	1031	2031	3052	4022	5029	4971	3989	3036	2052	1007	493	260	104	62	10	5	21.49
FYY NNNN	5	11	50	106	249	514	1012	2008	3009	3967	4950	5050	4035	3019	2027	1020	522	253	116	58	10	3	11.90
FXX NNON	5	9	46	89	217	419	912	1904	2860	3849	4839	5161	4176	3073	1962	915	394	168	55	21	4	3	76.19
FXX NNNN	8	15	52	89	228	455	920	1866	2807	3782	4831	5169	4115	2988	1944	913	407	189	53	26	4	2	70.70
FXY NNON	4	7	42	85	242	482	948	1891	2861	3900	4875	5125	4084	3083	2041	994	486	241	100	64	22	15	43.18
FXY NNNN	5	9	44	107	245	481	941	1915	2840	3830	4880	5120	4087	3038	2017	1003	479	238	97	54	11	9	30.40
FYX NNON	3	11	55	96	272	519	1014	1987	2985	3971	5046	4954	3997	3009	1991	1054	542	293	133	76	20	11	40.79
FYX NNNN	3	6	40	91	255	525	1000	1973	2985	3964	4977	5023	3999	3045	2019	1026	518	264	113	61	15	8	15.36
FYB NNON	6	9	56	106	262	554	1088	2126	3120	4110	5136	4854	3887	2928	1988	1057	550	291	141	82	24	11	54.80
FYB NNNN	7	14	54	115	277	551	1085	2159	3125	4126	5103	4897	3949	2966	2002	1053	578	320	144	75	22	8	59.11
FBY NNON	4	5	48	108	253	580	1171	2239	3349	4333	5259	4741	3822	2893	2001	1100	633	367	183	89	37	22	232.37
FBY NNNN	7	12	60	120	319	635	1212	2366	3401	4392	5340	4660	3772	2891	2050	1156	687	409	221	125	48	32	401.87
FXA NNON	2	4	40	84	222	461	966	2034	3019	4022	4974	5026	4076	3144	2162	1138	592	304	132	63	19	16	67.25
FXA NNNN	7	13	55	106	268	537	1024	2025	3038	3933	4901	5099	4119	3075	2065	1054	515	245	103	53	11	5	24.01
FAX NNON	6	9	46	92	264	552	1093	2125	3185	4188	5130	4870	3890	2993	2146	1275	831	606	412	326	143	68	2509.59
FAX NNNN	6	13	75	136	301	582	1129	2210	3245	4265	5249	4751	3832	2966	2115	1311	848	605	394	282	104	64	1640.78
FXB NNON	4	7	46	97	241	480	996	1987	2962	3969	4931	5069	4116	3143	2173	1140	611	335	158	91	41	27	139.93
FXB NNNN	5	12	53	97	256	506	1023	2014	2995	3909	4920	5080	4099	3091	2078	1063	548	284	114	61	15	10	21.42
FBX NNON	8	12	48	103	290	578	1118	2208	3214	4230	5179	4821	3872	3006	2156	1323	903	644	440	338	193	139	4536.61
FBX NNNN	4	10	59	142	323	638	1192	2293	3336	4331	5246	4754	3880	3015	2181	1327	907	639	414	308	142	100	2752.47
FYA NNON	13	21	65	124	296	556	1098	2127	3133	4135	5156	4844	3857	2907	1977	1045	541	297	137	83	28	16	78.08
FYA NNNN	11	19	49	97	266	569	1091	2152	3140	4116	5125	4875	3892	2993	2002	1060	583	316	145	76	22	12	76.72
FAY NNON	4	8	55	111	282	565	1111	2200	3236	4291	5271	4729	3799	2906	1981	1067	615	368	180	101	31	23	176.08
FAY NNNN	7	10	75	135	301	612	1173	2272	3370	4404	5349	4651	3722	2857	1998	1148	676	407	224	143	57	36	458.28
RXY 2N2N	771	883	1282	1516	1927	2338	2909	3720	4355	4889	5429	4571	4041	3453	2805	1994	1468	1114	735	547	253	190	135161.82
RX 2N2N	1003	1126	1539	1790	2204	2595	3086	3704	4165	4558	4971	5029	4657	4243	3713	3065	2592	2190	1770	1547	1122	987	410177.08
RY 2N2N	3	6	67	116	290	514	1005	2041	2993	3965	5025	4975	3922	2988	1980	980	485	222	93	53	10	4	40.85
FX NONN	3	7	46	82	220	444	909	1787	2779	3787	4795	5205	4148	3071	2025	920	427	168	55	21	4	3	80.35
FY NONN	6	13	57	111	259	501	1004	1990	2975	4013	5009	4991	4030	3015	2026	1006	506	262	104	47	10	4	8.34
FX 2NNN	2	8	52	102	233	470	923	1913	2890	3914	4936	5064	4062	3046	2010	951	423	207	75	24	4	2	38.60
FY 2NNN	5	11	49	101	239	523	986	2001	3019	4015	5006	4994	4036	3040	2032	1011	515	246	98	50	12	5	14.12

N= 64

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
FYY	NNON	6	14	61	115	253	509	1032	1989	3007	4030	5017	4983	3941	2940	1958	966	488	230	99	54	13	4	18.29
FYY	NNNN	4	5	47	90	268	507	995	1980	3000	4006	5050	4950	3955	2971	1978	966	497	269	106	52	10	6	19.35
FXX	NNON	3	8	48	98	244	482	977	1898	2875	3868	4913	5087	4071	3047	1991	960	476	230	78	39	7	5	21.73
FXX	NNNN	4	5	38	84	247	464	913	1873	2864	3893	4930	5070	4019	2976	1933	987	466	241	90	52	11	5	35.23
FXY	NNON	7	12	52	103	237	479	929	1933	2906	3973	5014	4986	4044	3018	2021	1014	489	241	100	51	11	4	21.62
FXY	NNNN	8	12	51	104	212	437	964	1937	2934	3928	4938	5062	4016	2992	1943	974	489	242	106	56	9	5	28.07
FYX	NNON	4	7	36	93	258	555	1061	2033	2978	3977	5000	5000	3990	2959	2009	1006	509	265	110	64	13	5	28.80
FYX	NNNN	5	10	48	96	250	490	992	2001	3011	4016	5008	4992	3994	3000	1996	1002	520	255	110	56	14	7	4.79
FYB	NNON	3	9	56	106	256	541	1080	2041	3076	4065	5125	4875	3868	2855	1925	985	531	295	126	75	18	7	45.98
FYB	NNNN	6	8	47	91	259	510	1059	2076	3136	4138	5153	4847	3853	2901	1938	1019	550	305	148	80	20	14	59.33
FBY	NNON	6	13	51	104	282	554	1128	2270	3342	4366	5348	4652	3664	2796	1987	1068	604	345	158	94	32	22	197.12
FBY	NNNN	3	4	51	111	295	608	1226	2367	3434	4473	5442	4558	3651	2844	1978	1133	635	392	215	137	52	30	427.56
FXA	NNON	10	20	69	126	270	533	1013	2004	2962	3919	5007	4993	3963	3038	2051	1074	579	310	145	69	19	9	60.60
FXA	NNNN	7	17	49	96	264	503	995	1976	3023	4024	5009	4991	3971	2999	2009	1043	550	280	127	70	13	4	28.95
FAX	NNON	7	12	68	122	272	542	1057	2167	3257	4237	5224	4776	3862	2969	2066	1173	691	431	253	161	69	53	664.20
FAX	NNNN	4	7	63	122	274	571	1156	2284	3358	4361	5309	4691	3789	2889	2055	1222	749	485	262	176	80	60	924.04
FXB	NNON	4	9	56	101	261	514	971	1968	2981	3998	4998	5002	3994	3015	2077	1087	611	321	146	74	16	7	44.74
FXB	NNNN	7	16	59	104	246	467	1001	2018	3049	4016	5004	4996	3989	2991	2010	1042	551	279	130	73	16	8	28.66
FBX	NNON	2	10	52	124	294	600	1172	2254	3287	4304	5260	4740	3820	2978	2062	1203	753	460	270	183	87	66	1031.54
FBX	NNNN	5	15	60	129	312	612	1220	2346	3423	4403	5318	4682	3780	2941	2078	1229	774	497	280	201	87	63	1095.57
FYA	NNON	6	8	52	101	273	544	1084	2092	3078	4099	5141	4859	3871	2882	1916	1026	536	292	134	73	19	10	42.89
FYA	NNNN	4	7	54	98	264	529	1084	2097	3141	4141	5154	4846	3894	2913	1951	1037	551	300	143	79	20	13	52.65
FAY	NNON	5	11	61	105	263	564	1115	2228	3336	4385	5336	4664	3735	2842	1975	1065	606	351	158	94	22	16	160.68
FAY	NNNN	8	11	57	121	274	598	1206	2351	3441	4472	5391	4609	3698	2820	1954	1107	642	410	208	141	44	28	392.31
RXY	2N2N	689	810	1218	1441	1925	2378	2930	3713	4317	4835	5383	4617	4064	3495	2889	2122	1546	1154	804	636	342	252	118293.21
RX	2N2N	1039	1194	1619	1889	2311	2707	3175	3819	4270	4702	5096	4904	4494	4080	3623	2996	2524	2085	1685	1444	1052	911	398083.55
RY	2N2N	5	7	48	101	253	496	1021	2040	3080	4090	5107	4893	3898	2897	1913	979	512	239	88	42	11	7	18.11
FX	NONN	6	10	42	87	231	471	960	1917	2876	3897	4939	5061	4043	3042	1976	951	471	220	83	38	6	2	20.45
FY	NONN	5	9	54	104	256	493	1027	2015	2989	3999	4979	5021	3968	2944	1936	989	502	280	109	44	14	8	26.83
FX	2NNN	4	10	53	87	227	451	940	1904	2901	3910	4906	5094	4088	3014	2008	966	480	219	89	43	15	8	28.13
FY	2NNN	2	6	46	100	254	496	1005	1995	3021	4036	5033	4967	4003	2998	1982	993	540	276	119	55	11	7	16.40

N= 128

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
FYY NNON	5	6	41	103	251	499	973	2004	2974	3988	4946	5054	4001	3011	1993	1008	479	257	87	39	8	1	28.97
FYY NNNN	2	8	51	115	248	479	979	1990	2975	3968	4953	5047	4014	3036	1978	999	484	233	85	40	5	2	19.72
FXX NNON	0	9	46	105	260	510	947	1992	2974	4010	4986	5014	3965	2967	1981	956	436	191	86	46	15	9	47.42
FXX NNNN	6	10	50	96	245	483	1000	2001	3018	4036	5009	4991	4002	3013	1970	933	448	232	95	47	9	6	14.06
FXY NNON	5	11	39	92	240	489	953	1948	2970	3965	5003	4997	3954	2971	1952	980	476	256	95	50	13	6	18.03
FXY NNNN	3	9	46	86	223	480	964	1949	3000	4025	5015	4985	4024	2971	1980	973	494	235	97	50	9	6	16.44
FYX NNON	9	13	54	102	275	554	1005	2008	2952	3978	4958	5042	4112	3062	2053	1018	484	235	94	50	10	6	32.10
FYX NNNN	9	14	65	123	277	533	1001	2021	2999	3955	4930	5070	4043	3025	2006	973	480	236	91	39	9	4	18.94
FYB NNON	7	13	58	103	259	514	1064	2085	3105	4061	5085	4915	3954	2921	1990	1051	526	283	132	72	14	7	35.06
FYB NNNN	7	15	56	113	274	558	1080	2121	3132	4119	5122	4878	3923	2921	2025	1023	542	294	125	62	13	6	34.38
FBY NNON	6	14	60	122	312	588	1112	2186	3242	4274	5288	4712	3790	2860	1950	1029	571	336	174	100	27	15	129.39
FBY NNNN	7	16	72	132	306	592	1213	2304	3367	4378	5337	4663	3749	2854	1922	1064	628	376	201	131	42	22	291.17
FXA NNON	5	10	55	114	269	502	965	2019	3137	4146	5127	4873	3911	2913	1965	1020	539	273	127	71	23	14	54.34
FXA NNNN	5	12	56	96	256	510	1060	2155	3170	4183	5118	4882	3919	2953	2008	1026	536	289	135	70	19	9	44.52
FAX NNON	5	6	55	112	305	597	1173	2234	3301	4339	5295	4705	3808	2869	1961	1079	644	398	214	124	46	31	324.97
FAX NNNN	4	14	57	122	328	622	1208	2375	3438	4450	5347	4653	3742	2832	1939	1142	719	425	232	148	61	35	539.33
FXB NNON	5	9	42	94	263	493	980	2052	3105	4125	5117	4883	3926	2986	2010	1027	560	291	123	78	26	12	56.50
FXB NNNN	3	5	47	101	252	500	1049	2120	3160	4109	5097	4903	3943	2950	2015	1048	549	288	134	66	19	8	41.49
FBX NNON	8	12	78	141	327	617	1161	2220	3301	4290	5215	4785	3830	2961	2028	1132	692	411	224	149	53	36	423.54
FBX NNNN	14	26	88	172	334	659	1232	2343	3389	4385	5363	4637	3755	2878	2019	1181	732	451	253	165	73	49	765.74
FYA NNON	6	12	48	113	278	536	1049	2095	3047	4059	5059	4941	3950	2930	1972	1055	535	278	134	75	16	6	38.17
FYA NNNN	7	11	59	112	297	568	1072	2098	3110	4116	5105	4895	3933	2968	1983	1031	534	300	132	64	12	7	33.00
FAY NNON	4	15	59	117	296	570	1128	2167	3250	4288	5276	4724	3737	2794	1901	997	576	332	177	107	38	21	179.62
FAY NNNN	5	11	58	114	298	589	1157	2277	3343	4401	5347	4653	3714	2819	1888	1076	620	356	211	142	55	37	411.58
RXY 2N2N	576	675	1066	1281	1692	2123	2678	3450	4055	4604	5162	4838	4325	3804	3158	2352	1783	1389	938	742	406	310	96001.88
RX 2N2N	966	1113	1514	1753	2141	2525	3016	3684	4126	4538	4921	5079	4685	4283	3730	3103	2616	2206	1811	1579	1143	1013	406794.43
RY 2N2N	6	10	47	95	239	485	993	1946	2930	3961	4969	5031	3980	2997	2016	1018	471	240	83	39	8	3	17.07
FX NONN	2	8	44	82	230	471	988	1935	2990	3984	5003	4997	3991	3003	1999	976	478	214	92	46	3	2	24.73
FY NONN	8	14	55	98	255	501	933	1986	3038	4022	4989	5011	4045	3042	1982	955	462	230	90	38	6	1	31.94
FX 2NNN	6	15	49	101	243	513	1046	2022	3015	4020	5063	4937	3987	3009	1961	980	497	247	109	52	9	4	19.68
FY 2NNN	8	9	32	84	232	500	953	1945	2991	4022	5060	4940	3991	2951	1973	971	479	244	93	36	7	4	33.36

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
FYY NNON	4	8	44	96	240	482	974	1999	3041	4066	5050	4950	3963	3020	2018	1007	486	244	99	46	14	11	19.40
FYY NNNN	3	8	44	84	244	474	976	2007	3030	4032	5025	4975	4043	3047	2027	1007	501	245	101	53	14	5	16.22
FXX NNON	7	8	46	100	252	514	1029	2025	2999	4012	5039	4961	3995	3013	2013	991	483	242	89	42	5	3	12.54
FXX NNNN	3	9	53	109	276	524	1018	2044	3035	4026	4984	5016	4030	3051	2043	987	484	237	85	42	5	2	14.46
FXY NNON	5	7	44	84	239	494	1017	2025	2970	4019	5023	4977	3976	2971	1961	1003	526	254	102	48	13	8	18.51
FXY NNNN	3	7	43	99	270	517	1055	2037	3046	4023	5004	4996	4008	3021	2016	1020	493	253	92	46	14	6	16.22
FYX NNON	4	10	51	92	233	478	972	1999	3011	4038	5012	4988	4009	3006	2027	989	498	249	97	50	11	9	12.77
FYX NNNN	5	10	49	91	230	478	954	1991	2987	3969	4996	5004	4046	3039	2028	1051	509	254	99	47	12	4	14.68
FYB NNON	2	8	46	100	270	525	1073	2132	3131	4155	5152	4848	3872	2932	1996	1006	528	286	121	73	23	15	49.34
FYB NNNN	7	15	53	96	245	541	1087	2097	3158	4124	5108	4892	3974	3009	2011	1039	556	297	134	85	24	14	64.63
FBY NNON	3	11	56	116	268	541	1106	2242	3362	4366	5303	4697	3780	2836	1948	1045	588	347	173	98	28	14	159.94
FBY NNNN	3	13	50	115	288	612	1205	2372	3427	4421	5334	4666	3706	2801	1962	1102	670	403	210	132	44	28	373.93
FXA NNON	8	12	56	101	254	520	1050	2108	3194	4129	5134	4866	3919	2925	1957	1054	542	292	128	70	18	10	47.35
FXA NNNN	4	7	48	107	301	578	1107	2220	3179	4137	5129	4871	3950	2985	2041	1049	564	279	131	69	13	9	65.15
FAX NNON	1	4	50	105	295	578	1152	2211	3306	4281	5261	4739	3759	2890	2002	1041	601	344	183	118	41	28	242.06
FAX NNNN	7	14	69	131	312	616	1208	2323	3379	4400	5399	4601	3701	2875	1979	1102	644	391	224	146	52	39	466.20
FXB NNON	3	6	43	85	230	510	1021	2086	3115	4090	5178	4822	3869	2919	1975	1057	545	285	127	73	19	14	57.65
FXB NNNN	5	9	38	89	260	556	1130	2189	3180	4145	5119	4881	3911	2984	2040	1061	552	294	134	72	13	8	55.66
FBX NNON	6	12	59	111	268	546	1099	2226	3304	4307	5248	4752	3003	2857	1972	1081	612	355	189	109	44	27	230.16
FBX NNNN	10	12	56	102	283	598	1239	2340	3406	4373	5327	4673	3737	2842	2010	1126	675	415	238	142	59	42	546.36
FYA NNON	4	8	49	92	254	538	1061	2096	3160	4154	5141	4859	3917	2931	1949	1021	533	284	123	73	22	16	51.16
FYA NNNN	3	10	45	102	246	530	1060	2113	3170	4101	5096	4904	3965	2993	2020	1081	568	294	132	84	23	13	62.17
FAY NNON	5	12	53	104	280	601	1150	2269	3318	4387	5314	4686	3767	2860	1946	1067	591	353	171	96	32	19	181.67
FAY NNNN	11	16	79	149	331	654	1233	2381	3462	4449	5409	4591	3736	2822	1944	1132	663	401	205	119	46	34	425.68
RXY 2N2N	566	697	1088	1314	1710	2159	2690	3428	4071	4624	5152	4848	4326	3805	3171	2383	1823	1429	1017	786	437	336	98983.63
RX 2N2N	1031	1192	1569	1822	2192	2548	3047	3650	4153	4588	4958	5042	4652	4248	3770	3094	2612	2238	1834	1560	1100	969	416848.70
RY 2N2N	5	13	67	134	292	548	1052	1976	2965	3988	4958	5043	4007	3026	2008	979	475	243	90	49	12	5	27.26
FX NONN	4	10	56	108	240	520	1037	2041	3075	4055	5049	4951	3992	3021	2010	993	480	239	101	42	7	2	17.92
FY NONN	2	5	47	90	230	480	987	2001	3036	4023	5030	4970	3996	3012	2046	997	488	243	103	47	8	6	14.25
FX 2NNN	7	13	57	103	237	494	1004	1964	3013	4030	5001	4999	4000	3002	2037	995	504	234	88	40	7	3	16.12
FY 2NNN	4	6	38	89	216	467	1001	1979	2998	3996	5031	4969	3995	2995	2025	987	490	260	103	49	10	6	17.25

N= 512

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
FYY NNON	4	8	62	104	257	478	1026	2047	3116	4100	5093	4902	3877	2901	1948	947	467	226	93	45	11	10	35.72
FYY NNNN	6	8	45	90	251	512	1061	2070	3095	4094	5082	4918	3920	3004	2013	957	432	238	89	37	12	10	46.25
FXX NNON	3	8	38	107	240	519	990	2001	2991	3917	4931	5069	4042	3037	2057	1019	505	268	103	54	12	7	29.47
FXX NNNN	2	7	57	99	245	503	1013	2017	2967	3945	4949	5051	4045	3033	2041	1009	521	258	102	50	16	7	16.61
FXY NNON	8	14	51	105	277	557	1015	2018	3022	4058	5075	4925	3929	2952	1955	962	472	239	95	61	12	7	24.65
FXY NNNN	3	13	60	110	247	516	996	2046	3052	4057	5079	4921	3956	2934	1942	992	490	228	97	44	11	3	24.71
FYX NNON	7	12	41	98	247	490	967	1967	2939	3981	4977	5023	4029	3012	2023	1000	489	258	108	58	18	11	19.29
FYX NNNN	2	5	48	100	230	489	1013	2029	3030	3987	5003	4997	4043	3082	2025	995	485	261	98	51	14	8	23.63
FYB NNON	7	10	51	99	250	505	1069	2092	3182	4199	5203	4797	3776	2832	1935	998	503	254	115	65	20	13	52.18
FYB NNNN	6	12	46	98	268	571	1089	2186	3208	4235	5209	4791	3853	2932	2002	1061	524	257	116	61	17	14	68.78
FBY NNON	5	12	56	102	281	563	1187	2336	3364	4416	5418	4582	3622	2740	1894	1026	572	333	157	89	22	14	176.26
FBY NNNN	3	11	62	125	334	657	1291	2414	3499	4511	5468	4532	3653	2782	1914	1060	613	371	188	118	40	27	350.64
FXA NNON	10	14	50	104	271	543	1055	2107	3082	4049	5076	4924	3923	2934	1974	1012	546	288	135	65	20	6	46.04
FXA NNNN	6	10	64	123	297	554	1089	2101	3105	4089	5102	4898	3911	2954	1958	1032	578	299	127	70	13	7	41.53
FAX NNON	7	11	49	102	282	534	1072	2189	3238	4279	5227	4773	3836	2909	1966	1148	658	398	209	141	42	31	322.97
FAX NNNN	5	9	62	117	287	590	1227	2358	3377	4325	5317	4683	3766	2846	2034	1193	693	451	253	166	56	41	595.27
FXB NNON	5	12	64	116	289	528	1075	2125	3133	4125	5129	4871	3859	2882	1919	999	537	283	131	72	17	7	40.02
FXB NNNN	5	14	62	118	274	538	1105	2177	3164	4139	5168	4832	3839	2915	1960	1018	580	295	123	69	12	7	57.71
FBX NNON	6	12	59	119	281	572	1108	2179	3282	4268	5230	4770	3844	2901	1984	1128	664	384	213	138	47	38	375.20
FBX NNNN	4	15	58	127	307	610	1194	2354	3389	4359	5301	4699	3787	2879	1996	1159	703	444	247	157	61	47	632.07
FYA NNON	4	10	48	100	251	510	1045	2053	3111	4119	5155	4845	3860	2904	1967	1024	508	264	121	70	20	13	34.63
FYA NNNN	4	6	53	102	262	558	1105	2172	3180	4153	5123	4877	3899	2981	2032	1064	535	265	123	69	18	14	57.26
FAY NNON	6	12	58	120	298	594	1138	2247	3362	4436	5389	4611	3655	2784	1903	1031	571	332	164	89	28	19	175.42
FAY NNNN	5	15	73	136	315	609	1204	2397	3499	4527	5521	4479	3611	2744	1917	1075	614	372	199	122	43	27	366.79
RXY 2N2N	506	630	993	1239	1634	2037	2636	3404	4004	4541	5070	4930	4392	3846	3199	2397	1831	1442	1052	860	469	339	87545.25
RX 2N2N	950	1105	1562	1773	2155	2554	3052	3687	4155	4611	4997	5003	4613	4196	3699	3034	2532	2144	1771	1513	1064	904	362484.89
RY 2N2N	4	10	49	93	228	485	1006	2007	2950	3970	5053	4948	3978	2957	1926	977	516	263	96	51	8	5	25.64
FX NONN	7	13	51	102	267	495	999	1938	2961	3963	4936	5064	4044	3059	2060	987	516	268	92	50	10	6	23.41
FY NONN	4	9	51	97	245	487	1012	2005	3018	4015	5015	4985	3993	2992	2013	979	472	231	81	34	8	3	10.34
FX 2NNN	3	4	40	89	245	510	999	2027	3006	3977	4992	5008	4052	3043	2024	1059	541	256	106	52	10	7	19.26
FY 2NNN	7	10	65	116	273	520	1059	2073	3089	4117	5069	4931	3978	2983	1985	1016	493	215	94	43	8	3	28.63

N=1024

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES										RIGHT-TAIL CUMULATIVE PROBABILITIES												
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
FYY NNON	5	13	50	92	253	489	965	1950	2975	4028	5060	4940	3912	2944	1950	980	499	256	101	55	9	3	16.78
FYY NNNN	1	6	39	84	217	451	943	2004	3017	4061	5057	4943	3977	2994	2003	1005	470	216	97	46	7	2	26.14
FXX NNON	4	5	53	98	258	527	1011	2017	2996	4047	5055	4945	3963	2981	1952	973	482	243	97	43	13	5	18.57
FXX NNNN	3	8	57	115	275	524	1012	2011	3020	4009	4988	5012	3968	2983	1975	999	497	252	81	42	6	2	16.38
FXY NNON	4	11	42	87	230	477	969	2009	3018	3994	5058	4942	3911	2936	1950	978	487	265	110	56	17	9	21.74
FXY NNNN	6	5	45	104	241	486	992	1987	3004	4021	5002	4998	4013	3026	1965	980	498	246	95	52	13	5	16.92
FYX NNON	2	3	38	90	216	464	995	2007	3066	4065	5097	4903	3949	2951	1997	993	474	235	93	54	9	3	26.06
FYX NNNN	6	11	45	94	236	457	961	2047	3064	4069	5081	4919	3982	2986	1999	982	470	225	85	36	7	3	22.34
FYB NNON	2	6	37	94	233	479	1011	2074	3122	4193	5209	4791	3782	2872	1963	1007	543	296	125	64	21	10	58.30
FYB NNNN	1	2	46	94	262	514	1058	2117	3213	4214	5206	4794	3839	2935	1994	1023	536	276	108	61	19	8	53.43
FBY NNON	7	8	53	110	279	569	1110	2191	3299	4317	5289	4711	3755	2818	1914	1057	581	333	158	106	35	21	169.52
FBY NNNN	2	5	56	119	290	564	1186	2337	3432	4427	5383	4617	3710	2803	1974	1103	639	374	200	120	46	34	373.22
FXA NNON	3	11	50	111	278	524	1057	2086	3121	4130	5152	4848	3859	2874	1922	992	547	295	123	69	19	10	39.71
FXA NNNN	4	12	52	120	294	554	1093	2135	3161	4154	5167	4833	3891	2964	1998	1046	560	277	131	64	19	9	50.03
FAX NNON	11	15	72	141	298	605	1156	2287	3355	4367	5322	4678	3695	2801	1941	1045	587	339	173	102	35	26	229.96
FAX NNNN	7	11	83	148	343	634	1223	2384	3458	4476	5418	4582	3656	2788	1961	1099	652	387	221	140	41	29	413.59
FXB NNON	3	10	61	110	275	545	1064	2102	3136	4138	5185	4815	3832	2861	1924	1042	540	301	137	75	21	11	54.08
FXB NNNN	7	17	56	122	286	546	1076	2140	3162	4178	5153	4847	3914	2950	2020	1046	550	297	132	66	18	8	46.36
FBX NNON	2	9	58	114	273	555	1129	2242	3362	4378	5364	4636	3698	2807	1923	1062	593	344	180	96	37	22	210.23
FBX NNNN	9	15	50	108	285	594	1235	2416	3517	4481	5419	4581	3655	2786	1934	1090	663	399	222	145	41	27	427.11
FYA NNON	0	6	47	85	237	503	1045	2108	3140	4175	5234	4766	3822	2889	1956	996	540	292	120	59	20	10	55.66
FYA NNNN	1	7	43	91	272	538	1065	2113	3225	4220	5184	4816	3874	2949	1996	1016	531	262	116	59	18	9	49.94
FAY NNON	9	11	59	99	263	563	1123	2179	3342	4341	5324	4676	3741	2832	1897	1048	577	338	179	93	35	22	203.96
FAY NNNN	7	13	56	119	304	635	1183	2320	3390	4419	5373	4627	3692	2832	1949	1105	644	384	216	119	45	27	341.35
RXY 2N2N	503	614	954	1182	1582	1999	2570	3360	3956	4470	4971	5029	4495	3927	3282	2509	1923	1491	1067	835	495	396	92447.50
RX 2N2N	914	1056	1483	1703	2149	2547	30	681	4140	4551	4956	5044	4642	4205	3730	3103	2587	2214	1794	1549	1095	960	367602.01
RY 2N2N	2	4	43	100	243	501	94	2015	2976	3968	4926	5075	4091	3093	2066	1013	512	251	93	46	9	3	15.18
FX NONN	4	10	59	123	276	538	1036	2036	3002	3974	4969	5031	4016	3014	1990	978	501	263	94	30	6	3	25.86
FY NONN	4	10	38	92	251	482	954	1983	3007	3991	4999	5001	3964	2976	1999	985	478	229	92	41	12	4	18.12
FX 2NNN	0	4	49	108	233	488	1024	2013	3020	4011	5014	4986	4002	3022	2040	989	500	265	92	37	10	2	31.14
FY 2NNN	4	8	45	101	246	477	988	1985	3018	4038	5013	4987	3971	2981	2034	999	501	231	86	43	7	1	16.27

N= 2

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX	1	0	0	0	0	0	0	0	37	2183	6821	8021	2001	1947	1922	1845	1520	1106	701	398	264	166	144	24099.65
ZY	1	5	12	48	88	220	439	965	1951	2928	3898	5126	5117	3915	2965	1948	983	497	232	110	52	11	3	44.44
ZX	2	0	0	0	0	0	0	0	36	2198	6738	8031	2012	1951	1929	1859	1546	1058	614	364	251	160	139	23105.07
ZY	2	7	15	53	100	227	464	987	1979	2930	3834	5045	5199	3906	3027	2037	995	447	219	109	55	11	7	92.42
ZX	3	0	0	0	0	0	0	0	32	2187	6811	8053	1969	1935	1914	1844	1528	1084	687	371	249	160	140	23833.63
ZY	3	3	7	50	100	235	430	933	1973	2947	3854	5104	5150	3873	2965	1980	994	484	246	111	66	5	2	88.68
ZX	4	0	0	0	0	0	0	0	44	2190	6751	8037	1989	1943	1918	1844	1535	1086	678	375	250	167	146	23650.42
ZY	4	8	17	59	121	257	488	970	1956	2903	3777	5048	5187	3962	3010	2024	986	484	242	96	58	13	5	75.00
ZX	1234	0	0	0	0	0	0	32	2473	4085	4348	5253	4874	3484	2424	1828	1157	661	454	309	221	127	102	6410.20
ZY	1234	16	25	59	108	223	453	923	1930	2876	3813	4981	5150	3937	3002	1975	985	472	255	133	85	22	13	123.36
TX	1	1273	1273	1273	1273	1307	2820	4450	7326	7868	7996	8021	2001	1967	1463	106	84	49	22	7	3	1	1	339269.65
TY	1	122	122	122	135	265	485	891	1923	2957	4053	5126	5117	4072	2983	1960	947	504	269	142	135	135	135	6335.24
TX	2	1283	1283	1283	1283	1328	2783	4408	7281	7861	7984	8031	2012	1960	1452	110	92	57	29	14	7	3	3	343657.50
TY	2	130	130	130	142	279	516	952	1903	2968	3964	5045	5199	4086	3052	2026	975	548	290	130	118	118	118	5860.97
TX	3	1307	1307	1307	1307	1355	2897	4493	7349	7904	8026	8053	1969	1940	1454	99	86	61	30	9	5	4	4	356834.83
TY	3	135	135	135	148	272	522	944	1986	2988	4038	5104	5150	4058	3016	1975	960	520	281	140	132	132	132	6796.24
TX	4	1300	1300	1300	1300	1354	2840	4465	7298	7861	7998	8037	1989	1954	1470	109	98	65	39	21	13	6	6	352599.69
TY	4	137	137	137	147	268	534	934	1985	2933	3971	5048	5187	4127	3069	2005	942	498	268	141	125	125	125	6571.07
TX	1234	2331	2957	3884	4027	4104	4113	4125	4161	4240	4464	5253	4874	3304	1943	1039	248	73	20	5	3	1	1	1183242.86
TY	1234	23	33	76	121	256	459	912	1893	2822	3862	4981	5150	4005	2964	1952	952	498	277	135	80	21	14	137.84
ZX	0	0	0	0	0	0	0	0	35	2224	6791	8067	1962	1919	1901	1832	1526	1083	679	398	280	168	146	
ZX		00	00	00	00	00	00	00	368	21964	67824	80418	19866	19390	19168	18448	15310	10834	6718	3812	2588	1642	1430	
ZY	0	8	12	49	105	237	468	961	1969	2976	3912	5163	5070	3891	2966	1974	1001	513	263	129	60	11	7	
ZY		62	126	518	1028	2352	4538	9632	19636	29368	38550	50972	51446	39094	29866	19926	9918	4850	2404	1110	582	102	46	
TX	0	1318	1318	1318	1318	1363	2960	4562	7346	7898	8032	8067	1962	1927	1454	124	98	62	28	13	7	6	6	
TX		12962	12962	12962	12962	13414	28580	44756	73200	78784	80072	80418	19866	19496	14586	1096	916	588	296	128	70	40	40	
TY	0	126	126	126	135	278	522	965	2007	3071	4080	5163	5070	4043	2991	2021	961	525	288	140	130	130	130	
TY		1300	1300	1300	1414	2724	5158	9312	19608	29834	40212	50972	51446	40772	30222	19974	9570	5190	2792	1386	1280	1280	1280	

N= 4

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX 1	0	0	0	0	0	0	0	693	4990	6494	6600	3406	3275	2913	2144	1054	643	491	346	261	116	91	16165.72
ZY 1	4	11	59	107	281	526	1054	2056	3044	4054	5092	5086	4062	3061	2091	1023	510	269	92	49	8	4	27.71
ZX 2	0	0	0	0	0	0	0	641	4998	6484	6568	3448	3326	2965	2155	1055	646	528	380	298	121	87	16816.71
ZY 2	3	11	55	107	264	513	1023	2014	3015	4035	5094	5081	3985	3013	2037	1026	496	271	108	57	11	3	15.40
ZX 3	0	0	0	0	0	0	0	625	4950	6508	6603	3413	3277	2956	2179	1052	644	484	349	261	116	84	16377.80
ZY 3	3	13	59	99	277	511	982	2052	3024	3999	4998	5171	4105	3083	2047	1019	530	258	87	45	8	3	40.58
ZX 4	0	0	0	0	0	0	0	635	4969	6470	6546	3461	3344	3002	2230	1072	621	489	337	251	110	89	16459.70
ZY 4	4	10	52	96	259	511	1028	2055	3063	4065	5103	5094	4020	3022	2024	991	527	268	94	45	6	3	17.50
ZX 1234	0	0	0	0	0	15	685	2031	3268	4619	5565	4501	3548	2613	1845	1038	617	400	222	155	71	54	1522.30
ZY 1234	10	14	56	101	256	483	954	1932	2950	3953	4995	[illegible]	3979	2990	1921	942	480	230	91	49	12	9	30.86
TX 1	277	484	1851	3112	4890	5927	6399	6536	6550	6563	6600	3406	3247	2008	506	40	16	1	1	1	0	0	122312.50
TY 1	6	7	60	105	261	511	1026	2033	3035	4022	5092	5086	3995	3025	2031	1012	512	259	110	57	15	7	13.62
TX 2	307	479	1786	3021	4838	5905	6386	6509	6523	6535	6568	3448	3280	1990	509	43	22	0	0	0	0	0	119123.06
TY 2	5	10	51	105	244	485	981	1999	2964	3977	5094	5081	3957	2986	2006	973	493	241	79	45	10	6	15.68
TX 3	314	523	1915	3088	4862	5936	6413	6544	6560	6566	6603	3413	3240	2009	486	41	18	1	1	1	0	0	124546.19
TY 3	8	13	60	115	268	529	1043	2025	2982	3949	4998	5171	4029	3034	2038	1027	520	268	111	52	8	4	15.31
TX 4	259	458	1871	3093	4879	5864	6357	6493	6507	6514	6546	3461	3308	2044	483	30	9	0	0	0	0	0	121031.34
TY 4	2	6	35	81	219	466	1008	2043	3035	4021	5103	5094	3975	2990	1946	979	486	244	111	56	10	4	19.98
TX 1234	1783	1790	1809	1826	1870	1952	2165	2844	3779	4709	5565	4501	3463	2241	1219	320	87	23	4	1	0	0	633333.90
TY 1234	8	17	56	93	235	463	940	1909	2941	3969	4995	5104	4001	2974	1913	957	457	238	102	49	11	4	30.83
ZX 0	0	0	0	0	0	0	0	638	5043	6531	6611	3398	3268	2912	2131	1047	655	507	369	258	115	91	
ZX	00	00	00	00	00	00	00	6464	49900	64974	65856	34252	32980	29496	21678	10560	6418	4998	3562	2658	1156	884	
ZY 0	1	8	48	92	263	512	1044	2100	3090	4101	5179	4992	3937	3008	1990	984	498	258	99	50	9	6	
ZY	30	106	546	1002	2686	5146	10262	20554	30490	40508	50932	50848	40218	30374	20394	10086	5122	2608	960	492	84	38	
TX 0	307	506	1864	3103	4965	5985	6442	6554	6572	6580	6611	3398	3237	1985	509	41	20	0	0	0	0	0	
TX	2928	4898	18614	30834	49868	59234	63904	65276	65424	65516	65856	34252	32624	20072	4986	390	170	94	04	04	00	00	
TY 0	4	10	43	95	260	506	1011	2041	3031	4113	5179	4992	3904	2821	1966	980	504	257	93	60	12	5	
TY	50	88	498	1002	2504	4998	10138	[illegible]	30094	40164	50932	50848	39720	29912	19964	9942	5030	2538	1008	540	110	52	

N= 8

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES										RIGHT-TAIL CUMULATIVE PROBABILITIES												
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX 1	0	0	0	0	0	0	22	2615	4215	4489	5336	4799	3490	2497	1885	1220	682	426	284	195	101	68	5808.83
ZY 1	4	8	46	96	237	494	1001	2010	3001	3938	5033	5101	4010	3056	2018	1001	483	246	96	45	10	6	13.37
ZX 2	0	0	0	0	0	0	29	2620	4226	4508	5386	4746	3387	2434	1826	1192	691	443	300	212	92	66	5790.37
ZY 2	5	9	56	119	268	502	1000	2026	3048	3933	5013	5098	3994	3045	1997	1034	495	247	93	46	13	9	37.42
ZX 3	0	0	0	0	0	0	20	2682	4334	4579	5427	4706	3373	2411	1787	1165	668	455	290	209	100	63	6145.48
ZY 3	3	5	42	100	235	482	1027	2036	3029	3968	5089	5034	3872	2957	1983	979	469	238	80	45	9	5	42.27
ZX 4	0	0	0	0	0	0	30	2581	4256	4525	5357	4770	3440	2475	1884	1234	695	453	289	196	102	70	5850.29
ZY 4	6	10	61	118	251	507	1016	1987	2997	3920	5027	5094	3931	3003	2039	1025	480	242	100	49	11	5	38.22
ZX 1234	C	0	0	0	32	273	776	2009	3199	4300	5377	4687	3682	2758	1850	1030	635	391	225	144	64	45	783.91
ZY 1234	2	8	56	96	235	484	991	1988	2958	3987	5029	5028	3932	2943	1995	1005	497	236	99	58	12	1	30.58
TX 1	2496	3098	3984	4142	4217	4227	4239	4271	4367	4609	5336	4799	3304	2016	1067	228	39	2	1	1	0	0	1334528.23
TY 1	2	8	49	107	237	462	987	1950	2978	3966	5003	5101	4046	3038	2003	991	503	227	82	39	6	5	22.25
TX 2	2435	3074	3983	4160	4233	4248	4260	4291	4380	4639	5386	4746	3217	1943	1070	241	43	5	0	0	0	0	1284666.36
TY 2	8	15	54	109	271	504	987	1978	2995	3985	5013	5098	4062	3044	1988	1025	499	243	102	52	9	7	17.12
TX 3	2586	3229	4125	4267	4321	4330	4337	4373	4459	4711	542·	4706	3197	1918	1034	227	34	4	1	0	0	0	1436383.03
TY 3	7	11	56	114	263	522	1024	2046	3000	4021	5089	5034	3928	2946	1971	976	495	227	77	42	6	3	17.80
TX 4	2462	3099	4009	4157	4243	4255	4265	4297	4391	4655	5357	4770	3297	2018	1124	227	36	5	0	0	0	0	1310423.81
TY 4	4	9	50	96	237	500	1021	1989	2961	3973	5027	5094	3991	2994	2013	1006	511	252	92	45	8	5	9.11
TX 1234	413	457	617	767	1052	1359	1808	2647	3496	4402	5377	4687	3638	2507	1388	494	160	41	4	0	0	0	34701.55
TY 1234	6	7	41	91	247	479	998	1963	2976	4010	5029	5028	3960	2942	1980	1004	494	238	88	51	10	3	18.49
ZX 0	0	0	0	0	0	0	28	2632	4286	4542	5395	4738	3418	2450	1824	1192	715	469	293	203	97	65	
ZX	00	00	00	00	00	00	258	26260	42634	45286	53802	47518	34216	24534	18412	12006	6902	4492	2912	2030	964	664	
ZY 0	5	10	58	116	260	497	978	1971	2973	3886	4998	5135	3955	2974	1988	986	480	257	101	49	6	1	
ZY	46	84	526	1098	2502	4984	10044	20060	30098	39290	50260	50924	3952·	30070	20050	10050	4814	2460	940	468	98	52	
TX 0	2533	3135	4051	4204	4278	4290	4302	4335	4418	4670	5395	4738	3262	1937	1048	234	37	1	0	0	0	0	
TX	25024	31270	40304	41860	42584	42700	42806	43134	44030	46568	53802	47518	32554	19764	10686	2314	382	34	04	02	00	00	
TY 0	8	15	55	97	256	524	1024	1952	2935	3930	4998	5135	4018	2970	2003	980	464	236	96	46	7	6	
TY	58	116	528	1046	2528	5024	10086	19630	29738	39750	50260	50924	40090	29984	19956	9956	4944	2370	898	448	72	52	

N= 16

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX 1	0	0	0	0	0	10	797	2150	3301	4660	5610	4456	3547	2630	1914	1118	678	421	219	152	59	48	1551.28
ZY 1	0	4	53	95	266	502	1018	2058	3089	4053	5112	4965	3935	2954	1970	948	514	254	100	47	6	3	29.22
ZX 2	0	0	0	0	0	12	792	2144	3365	4721	5621	4448	3533	2602	1834	1031	654	428	241	166	74	53	1696.41
ZY 2	4	9	48	97	260	529	1008	2039	3043	4082	5152	4944	3871	2936	1957	968	509	256	103	52	6	3	18.07
ZX 3	0	0	0	0	0	13	755	2099	3280	4588	5506	4562	3673	2655	1897	1076	683	434	220	156	69	44	1456.73
ZY 3	11	16	55	101	267	512	1026	2054	3044	3993	5024	5053	3989	2968	1993	998	485	252	113	49	9	5	21.45
ZX 4	0	0	0	0	(	10	786	2139	3341	4606	5543	4545	3576	2665	1928	1126	673	432	246	159	77	50	1588.42
ZY 4	4	7	54	98	256	500	1016	2015	2959	3883	4993	5114	4035	2999	2040	993	494	243	98	43	14	7	27.27
ZX 1234	0	0	2	23	106	359	927	2107	3250	4328	5355	4695	3724	2768	1937	1067	576	348	171	114	37	26	348.94
ZY 1234	10	16	62	127	268	528	1049	2009	3049	4069	5056	4988	3953	2891	1936	955	506	255	115	63	13	4	35.50
TX 1	1904	1906	1922	1941	1984	2048	2287	2927	3826	4752	5610	4456	3453	2293	1302	348	79	14	2	0	0	0	722297.81
TY 1	3	6	51	101	240	496	1004	2035	3073	4071	5112	4965	3938	2950	1959	983	510	253	104	46	10	3	11.59
TX 2	1913	1917	1930	1950	1997	2081	2304	2926	3906	4821	5621	4448	3417	2259	1196	356	81	14	0	0	0	0	729182.17
TY 2	5	11	47	101	259	522	1023	2014	3044	4086	5152	4944	3901	2941	1947	990	515	249	100	50	5	2	14.11
TX 3	1810	1819	1836	1855	1910	1997	2219	2898	3786	4691	5506	4562	3580	2286	1262	338	76	14	0	0	0	0	652707.66
TY 3	8	13	63	111	263	510	1033	2055	3033	3997	5024	5053	3993	2952	1970	999	500	236	100	52	11	6	13.60
TX 4	1892	1894	1914	1926	1975	2060	2270	2938	3827	4709	5543	4545	3493	2299	1310	349	89	19	3	2	0	0	713218.10
TY 4	6	8	48	98	245	512	1006	2000	2967	3904	4993	5114	4026	3008	1998	1018	485	236	92	41	12	6	17.52
TX 1234	193	237	405	523	780	1096	1607	2559	3481	4359	5355	4695	3672	2610	1570	568	203	79	15	1	0	0	8221.84
TY 1234	10	17	65	125	267	521	1068	2002	3045	4053	5056	4988	3944	2913	1936	937	498	256	115	55	10	6	34.18
ZX 0	0	0	0	0	0	10	808	2075	3314	4672	5579	4496	3637	2700	1962	1103	708	459	261	182	68	42	
ZX	00	00	00	00	00	110	7876	21214	33202	46494	55718	45014	35932	26504	19070	10908	6792	4348	2374	1630	694	474	
ZY 0	2	10	40	90	238	488	1014	2023	3009	3961	5020	5064	3948	3022	2044	1027	539	255	96	57	11	7	
ZY	42	92	500	962	2574	5062	10164	20378	30288	39944	50602	50280	39656	29758	20008	9868	5082	2520	1026	496	92	50	
TX 0	1842	1847	1860	1870	1914	2015	2248	2891	3779	4774	5579	4496	3536	2353	1304	358	76	12	2	0	0	0	
TX	18722	18766	18924	19084	19660	20402	22656	29160	38248	47494	55718	45014	34958	22980	12748	3498	802	146	14	04	00	00	
TY 0	5	9	46	91	248	507	998	2020	3030	3991	5020	5064	3974	3012	2042	985	512	267	102	51	6	3	
TY	54	94	510	1004	2510	5094	10128	20248	30294	40098	50602	50280	39664	29726	19832	9960	5044	2482	996	480	88	40	

N= 32

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX 1	0	0	0	0	45	301	866	2094	3290	4347	5456	4607	3623	2704	1888	1044	580	348	184	108	38	26	487.97
ZY 1	6	12	51	93	236	492	979	1987	3037	3979	5074	5000	3953	2945	1963	981	488	225	99	50	6	1	20.85
ZX 2	0	0	0	0	33	263	845	2053	3264	4323	5419	4650	3608	2740	1896	1060	613	358	192	118	55	32	612.84
ZY 2	9	16	49	92	247	508	1004	2010	3050	4031	5118	4957	3908	2914	1944	982	498	247	87	46	10	1	23.45
ZX 3	0	0	0	0	38	297	864	2118	3295	4368	5452	4598	3593	2719	1868	1041	619	377	211	131	42	25	566.95
ZY 3	5	12	60	110	249	509	1016	2037	3045	4021	5058	5005	3969	3023	1968	942	472	249	87	44	10	7	20.61
ZX 4	0	0	0	0	41	311	874	2104	3336	4417	5444	4617	3602	2746	1882	1035	609	360	184	113	41	25	519.01
ZY 4	7	11	65	126	276	532	1045	2074	3090	4083	5091	4969	3905	2936	1963	966	465	229	80	41	6	3	18.72
ZX 1234	0	0	17	47	172	406	988	2062	3140	4255	5320	4707	3667	2738	1850	973	533	304	151	93	29	15	171.62
ZY 1234	1	4	45	99	236	502	1042	2026	3070	4094	5141	4887	3820	2855	1887	950	466	221	93	52	15	7	30.86
TX 1	471	512	705	863	1148	1458	1910	2735	3601	4434	5456	4607	3570	2472	1433	441	118	26	5	0	0	0	45101.15
TY 1	4	10	49	104	252	492	996	2011	3050	4024	5074	5000	3980	2954	1960	985	497	238	92	46	2	1	13.44
TX 2	430	477	649	817	1130	1438	1881	2713	3558	4427	5419	4650	3586	2513	1437	479	115	30	3	0	0	0	37814.85
TY 2	10	12	50	90	256	504	993	2023	3057	4070	5118	4957	3933	2939	1966	984	495	244	89	46	6	2	17.86
TX 3	474	524	718	860	1167	1464	1921	2739	3609	4467	5452	4598	3536	2478	1406	468	129	29	5	0	0	0	46444.77
TY 3	6	15	59	100	249	513	1011	2030	3069	4047	5058	5005	3975	3035	1964	953	465	238	87	47	12	3	26.80
TX 4	491	529	723	862	1141	1467	1940	2772	3628	4524	5444	4617	3571	2499	1409	450	121	25	1	0	0	0	48788.22
TY 4	6	12	62	112	283	531	1061	2098	3094	4097	5091	4969	3940	2957	1982	966	498	212	96	47	10	4	27.10
TX 1234	93	127	256	344	604	913	1402	2344	3296	4298	5320	4707	3649	2597	1591	656	248	95	23	8	0	0	2208.65
TY 1234	1	8	41	104	239	480	1045	2034	3072	4100	5141	4887	3826	2868	1866	946	475	223	85	55	14	7	44.62
ZX 0	0	0	0	0	39	299	822	2105	3269	4329	5359	4714	3728	2826	1984	1122	651	374	201	126	55	44	
ZX	00	00	00	00	392	2942	8542	20948	32908	43568	54260	46372	36308	27470	19036	10604	6144	3634	1944	1192	462	304	
ZY 0	5	13	52	101	247	511	1025	2014	2989	3948	4979	5077	3992	2996	2027	1045	520	270	103	44	11	7	
ZY	64	128	554	1044	2510	5104	10138	20244	30422	40124	50640	50016	39454	29628	19730	9832	4886	2440	912	450	86	38	
TX 0	474	511	678	820	1098	1434	1918	2753	3548	4431	5359	4714	3692	2584	1509	497	119	29	7	3	0	0	
TX	4688	5106	6946	8444	11368	14522	19140	27424	35888	44566	54260	46372	35910	25092	14388	4670	1204	278	42	06	00	00	
TY 0	8	11	52	98	257	496	1027	1998	2991	3965	4979	5077	4025	3020	2028	1046	517	273	102	43	10	4	
TY	68	120	544	1008	2594	5070	10176	20320	30522	40406	50640	50016	39706	29810	19800	9868	4944	2410	932	458	80	28	

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES												RIGHT-TAIL CUMULATIVE PROBABILITIES												
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX	1	0	0	7	19	118	357	909	2069	3209	4298	5354	4687	3731	2778	1922	1077	614	362	166	97	35	23	305.63
ZY	1	3	10	46	99	244	462	995	2024	3001	3989	5013	5024	4025	2991	1977	1015	496	249	89	36	9	2	21.77
ZX	2	0	0	1	26	126	367	940	2128	3287	4392	5365	4682	3694	2736	1886	1023	571	349	185	109	40	28	365.86
ZY	2	3	7	55	100	243	515	1059	2072	3033	4043	5048	4997	3977	2955	1965	966	483	235	102	55	12	6	15.25
ZX	3	0	0	2	20	130	351	888	2048	3175	4316	5296	4743	3767	2823	1946	1068	616	348	187	103	34	19	278.02
ZY	3	6	13	61	118	259	514	988	1950	2956	4015	5044	4997	3958	2898	1938	995	483	240	103	57	14	6	23.55
ZX	4	0	0	4	20	98	341	898	2064	3201	4275	5276	4779	3821	2859	1978	1114	620	359	191	115	41	27	356.46
ZY	4	5	6	30	96	253	488	988	1990	2973	4024	5068	4981	3980	2944	1999	1016	505	255	117	61	13	6	28.91
ZX	1234	0	2	15	60	192	448	993	2070	3126	4195	5211	4820	3806	2868	1933	1041	552	317	149	94	27	17	121.27
ZY	1234	4	5	42	102	242	505	1018	2016	3023	4083	5128	4898	3897	2912	1957	1016	523	264	103	56	11	4	20.89
TX	1	205	243	416	530	757	1087	1609	2530	3426	4344	5354	4687	3689	2632	1575	602	189	55	9	0	0	0	9064.35
TY	1	4	11	46	99	252	469	1002	2039	3028	3986	5013	5024	4020	3024	1968	1014	469	243	96	45	7	4	19.70
TX	2	216	258	433	555	802	1114	1633	2587	3494	4429	5365	4682	3652	2568	1527	563	202	65	14	3	0	0	10078.55
TY	2	3	10	50	98	246	524	1029	2068	3049	4042	5048	4997	3966	2983	1961	966	484	242	96	49	12	7	10.07
TX	3	203	251	411	527	764	1051	1585	2479	3387	4348	5296	4743	3715	2656	1610	599	203	57	9	2	0	0	8965.38
TY	3	8	17	59	110	270	507	986	1934	2992	4014	5044	4997	3943	2909	1949	991	500	243	98	53	13	3	26.45
TX	4	175	211	390	506	764	1066	1570	2534	3419	4318	5276	4779	3765	2695	1615	607	205	64	15	7	0	0	6879.29
TY	4	4	9	36	90	229	476	988	2005	2994	4007	5068	4981	3965	2961	1997	1032	503	265	105	54	11	4	14.57
TX	1234	46	61	159	252	469	733	1332	2285	3222	4222	5211	4820	3779	2784	1733	802	332	129	40	18	1	0	616.86
TY	1234	1	6	40	94	240	508	1006	1997	3037	4084	5128	4898	3894	2921	1952	1018	527	260	101	56	10	7	21.21
ZX	0	0	0	8	28	115	351	889	2103	3183	4341	5376	4669	3734	2791	1921	1082	579	338	169	118	35	24	
ZX		00	00	44	226	1174	3534	9048	20824	32110	43244	53334	47120	37494	27974	19346	10728	6000	3512	1756	1084	370	242	
ZY	0	6	11	51	103	247	478	987	1963	2984	4021	5041	4997	4009	2960	1975	1033	521	255	95	46	7	4	
ZY		46	94	486	1032	2492	4914	10034	20036	29894	40184	50428	49992	39898	29496	19708	10050	4976	2468	1012	510	110	48	
TX	0	190	239	381	516	768	1073	1598	2514	3449	4391	5376	4669	3671	2632	1601	601	190	65	9	0	0	0	
TX		1978	2404	4062	5268	7710	10782	15990	25288	34350	43660	53334	47120	36984	26370	15856	5944	1978	612	112	24	0	00	
TY	0	4	9	35	96	244	498	980	2023	2986	4007	5041	4997	3987	2908	1988	1029	529	253	92	45	8	6	
TY		48	112	492	986	2482	4948	9970	20138	30086	40112	50428	49992	39762	29730	19726	10068	5010	2492	974	492	102	48	

N= 128

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX 1	0	0	13	44	158	404	925	2049	3089	4083	5156	4878	3828	2942	2043	1069	569	319	149	84	20	11	122.11
ZY 1	2	7	45	93	231	477	995	2017	2989	3968	4953	5087	4111	3063	2072	1023	499	237	102	55	9	4	17.82
ZX 2	1	2	16	44	146	409	901	1990	3066	4125	5142	4902	3900	2955	2064	1112	620	324	146	80	25	10	133.52
ZY 2	4	9	52	96	256	509	998	1979	2936	3899	4920	5127	4069	3065	2034	1026	515	252	109	57	11	7	11.13
ZX 3	1	1	15	34	138	377	903	2010	3090	4143	5209	4820	3839	2926	2037	1088	610	346	151	85	28	15	167.58
ZY 3	7	10	39	79	227	454	955	1981	2942	3965	5029	5004	4022	3013	2062	1032	516	264	118	57	7	3	24.88
ZX 4	0	0	10	31	160	427	960	2106	3135	4164	5225	4807	3773	2840	1914	1029	559	309	162	88	29	14	162.74
ZY 4	7	13	44	89	238	478	994	2039	3056	4072	5077	4952	3953	2983	2010	1007	502	242	105	53	7	3	12.54
ZX 1234	1	2	30	58	215	456	945	1968	3024	4069	5048	4965	3969	3013	2036	1093	598	302	146	74	21	14	72.50
ZY 1234	6	12	49	94	251	485	986	1980	2957	3915	4918	5094	4089	3082	2083	1018	526	265	116	53	9	3	16.09
TX 1	69	113	248	347	573	861	1381	2301	3224	4144	5156	4878	3807	2826	1761	705	265	84	18	4	0	0	1973.12
TY 1	3	6	41	93	234	465	976	2034	2978	3979	4953	5087	4113	3065	2059	1034	503	246	111	52	10	6	23.97
TX 2	95	113	235	365	568	850	1352	2283	3206	4164	5142	4902	3885	2816	1807	753	262	89	17	3	0	0	2129.49
TY 2	6	10	45	96	240	507	997	1977	2943	3918	4920	5127	4078	3073	2024	1026	509	258	112	50	12	5	12.83
TX 3	73	101	232	323	556	857	1358	2311	3216	4209	5209	4820	3809	2814	1758	735	285	90	20	1	1	0	1469.96
TY 3	4	8	43	85	222	443	956	1944	2970	3983	5029	5004	4034	3016	2036	1032	514	269	115	52	5	3	19.95
TX 4	86	115	251	362	602	908	1452	2380	3272	4213	5225	4807	3750	2727	1670	684	257	90	23	1	0	0	1976.48
TY 4	4	10	47	93	225	496	989	2021	3069	4078	5077	4952	3965	3006	2002	994	514	244	106	50	6	0	20.81
TX 1234	21	32	113	201	396	667	1153	2118	3090	4091	5048	4965	3953	2943	1910	918	378	174	49	22	2	1	193.52
TY 1234	7	13	51	95	258	488	979	1985	2962	3922	4918	5094	4094	3081	2081	1007	525	261	116	53	9	3	19.03
ZX 0	1	3	18	45	157	401	965	2028	3117	4189	5194	4840	3801	2846	1917	1011	519	289	142	85	27	19	
ZX	06	12	144	396	1518	4036	9308	20366	30994	41408	51852	48494	38282	29018	19960	10618	5754	3174	1500	844	258	138	
ZY 0	6	11	52	94	242	496	983	2013	3019	4024	5049	4983	3959	2977	2019	1008	494	245	81	42	8	4	
ZY	52	100	464	902	2388	4828	9870	20058	29884	39856	50056	50306	40228	30202	20394	10192	5052	2480	1030	528	84	42	
TX 0	86	117	241	357	589	886	1407	2327	3273	4233	5194	4840	3777	2724	1655	638	235	83	19	6	0	0	
TX	856	1118	2414	3508	5776	8724	13900	23204	32382	41926	51852	48494	38056	27814	17302	7030	2608	872	194	30	02	00	
TY 0	4	10	51	91	247	501	994	2016	3018	4050	5049	4983	3972	2980	2017	1008	488	230	84	42	7	5	
TY	42	88	454	916	2336	4824	9824	19984	29956	40016	50056	50306	40324	30280	20276	10188	5056	2494	1056	492	80	38	

N= 256

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX 1	1	1	22	62	181	416	972	2046	3110	4164	5168	4851	3865	2903	1975	1045	551	325	164	90	27	16	116.73
ZY 1	3	4	43	97	234	487	1013	2054	3045	4055	5072	4950	3944	2987	1982	986	496	259	104	58	13	7	13.66
ZX 2	2	4	19	50	183	397	936	1987	3028	4115	5122	4895	3924	2960	2043	1062	563	304	130	77	21	14	86.15
ZY 2	4	9	40	92	221	490	1011	2010	2955	3998	4996	5026	4015	3000	2022	1026	492	255	110	56	11	6	17.64
ZX 3	0	0	21	55	187	415	914	2016	3074	4175	5168	4853	3894	2894	1999	1025	555	315	155	94	23	13	113.98
ZY 3	8	11	40	102	248	501	1045	2067	3033	4046	5051	4961	4010	3008	2017	980	489	244	94	55	15	7	24.39
ZX 4	2	4	22	54	178	407	938	2034	3105	4106	5130	4893	3888	2965	1986	1066	578	312	141	79	28	17	103.76
ZY 4	5	10	40	96	253	485	1019	2019	2954	3933	4966	5052	4001	3028	2016	1022	503	267	100	53	5	5	25.15
ZX 1234	2	5	35	77	213	449	916	1974	3001	4008	5015	4997	3925	2921	1962	1048	565	309	148	75	24	17	72.97
ZY 1234	3	7	47	92	259	494	1007	2026	3003	3978	5006	5005	4002	2986	1962	1015	514	250	96	47	7	4	11.15
TX 1	43	66	152	242	451	753	1280	2231	3230	4187	5169	4851	3857	2814	1774	754	329	142	36	20	3	2	564.46
TY 1	4	8	46	101	231	484	1002	2041	3056	4055	5072	4950	3942	2996	1973	974	491	265	98	52	11	5	14.94
TX 2	37	56	158	237	412	718	1236	2189	3139	4136	5122	4895	3902	2881	1839	801	325	118	36	14	1	0	464.87
TY 2	3	7	41	88	228	493	1005	2022	2956	3991	4996	5026	4021	3013	2015	1011	491	253	113	59	11	4	14.92
TX 3	38	61	167	248	446	706	1216	2211	3194	4194	5168	4853	3865	2807	1830	800	330	137	41	13	1	1	505.88
TY 3	8	12	42	96	243	488	1036	2054	3086	4047	5051	4961	4008	3020	2012	969	485	244	89	56	13	6	25.33
TX 4	41	66	153	238	447	732	1236	2218	3211	4131	5130	4893	3868	2897	1804	802	335	130	35	10	2	0	539.01
TY 4	4	11	41	92	263	484	1013	2002	2963	3939	4966	5052	3996	3050	2070	1024	510	254	99	47	7	5	22.67
TX 1234	15	27	92	151	341	585	1073	2073	3056	4016	5015	4997	3906	2866	1871	895	438	212	72	35	7	3	84.16
TY 1234	4	9	46	92	252	491	1004	2028	3017	3994	5006	5005	3997	2992	1955	1009	511	253	96	46	6	4	10.41
ZX 0	0	4	25	59	168	396	938	2034	3061	4182	5187	4831	3831	2862	1958	1043	565	304	130	73	20	9	
ZX	10	26	218	560	1794	4062	9396	20234	30736	41484	51550	48646	38804	29168	19922	10482	5624	3120	1440	826	238	138	
ZY 0	4	8	45	96	248	496	978	2033	3032	4032	5073	4944	3937	2998	2025	1013	528	265	100	50	5	2	
ZY	48	84	416	964	2408	4918	10130	20366	30138	40128	50316	49666	39814	30042	20124	10054	5016	2580	1016	544	104	54	
TX 0	47	67	149	237	417	720	1245	2221	3163	4192	5187	4831	3815	2766	1764	783	320	118	28	7	0	0	
TX	412	632	1558	2404	4346	7258	12426	22140	31854	41684	51550	48646	38614	28330	18022	7880	3278	1290	352	128	14	06	
TY 0	5	8	45	92	252	502	985	2026	3036	4035	5073	4944	3942	3036	2030	1015	534	267	103	56	7	1	
TY	48	92	430	938	2434	4802	10082	20290	30194	40134	50316	49666	39818	30168	20100	9986	5022	2566	1004	540	96	42	

N= 512

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

	LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
	0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX 1	1	2	29	79	221	450	1005	2052	3068	4132	5162	4853	3883	2935	1937	1006	520	262	120	56	16	11	47.00
ZY 1	5	10	46	80	266	521	1027	2062	3043	3991	5008	5013	3986	2994	2006	962	471	250	82	40	6	5	32.15
ZX 2	1	2	35	78	215	433	927	1975	3011	4084	5082	4936	3927	2936	1946	1013	551	274	131	63	17	10	46.93
ZY 2	10	12	42	95	248	481	948	1947	2928	3925	4960	5058	3977	2969	1977	1035	507	258	92	49	11	5	27.16
ZX 3	2	5	29	65	191	446	967	1985	2980	3977	5015	5011	3981	2965	1984	1036	550	273	113	66	16	9	33.10
ZY 3	4	14	41	93	206	459	970	1968	2975	3967	4964	5056	4059	3048	2049	1016	509	258	111	58	13	9	24.94
ZX 4	4	7	43	87	223	487	1029	2067	3087	4126	5093	4927	3938	2971	1976	1042	569	310	122	72	21	12	46.30
ZY 4	3	8	53	94	253	498	987	1986	2991	4026	5053	4964	3986	2969	1963	955	476	219	82	46	10	6	13.76
ZX 1234	3	9	43	90	224	456	970	1966	3004	4000	5037	4972	3974	2952	1928	997	493	254	105	54	12	5	15.31
ZY 1234	7	14	55	109	248	483	998	1981	3004	3994	5019	4992	4011	2991	2050	1022	527	259	112	59	13	6	14.15
TX 1	20	37	130	205	380	672	1229	2200	3149	4168	5162	4853	3878	2864	1809	821	327	130	38	16	0	0	256.47
TY 1	5	11	49	84	260	510	1039	2060	3056	4002	5008	5013	3995	3009	2015	978	483	253	85	42	7	5	23.25
TX 2	24	49	128	208	372	628	1143	2136	3097	4099	5082	4936	3915	2882	1818	846	372	152	48	19	4	2	265.59
TY 2	9	14	40	94	244	480	955	1939	2937	3927	4960	5058	3992	2976	1984	1046	509	250	95	47	11	5	23.73
TX 3	26	39	114	186	381	666	1177	2113	3040	3997	5015	5011	3969	2996	1835	854	355	137	44	15	0	0	241.70
TY 3	5	14	40	93	210	458	972	1970	2971	3976	4964	5056	4064	3044	2039	1032	511	258	113	56	11	8	22.55
TX 4	32	55	133	216	433	730	1260	2220	3178	4155	5093	4927	3927	2916	1825	854	399	153	46	22	2	2	374.22
TY 4	3	10	52	92	259	502	976	1996	2989	4027	5053	4964	3998	2976	1958	964	482	218	82	45	8	3	19.17
TX 1234	11	21	86	136	316	568	1056	2052	3034	4009	5037	4972	3959	2922	1865	899	405	189	62	30	3	1	65.42
TY 1234	6	15	51	106	250	484	1002	1991	2999	3992	5019	4992	4002	3006	2059	1024	532	268	111	59	14	4	18.34
ZX 0	1	4	26	72	192	434	961	2013	3071	4130	5155	4861	3939	2991	1995	997	546	270	121	71	15	5	
ZX	18	40	324	762	2084	4500	9782	20184	30434	40898	51014	49176	39336	29596	19676	10188	5472	2778	1214	656	170	94	
ZY 0	5	7	43	105	259	516	1019	2014	2978	4010	5035	4982	3939	2969	1970	964	468	241	98	57	13	3	
ZY	54	102	450	934	2464	4960	9902	19954	29830	39838	50040	50146	39894	29898	19930	9864	4862	2452	930	509	106	56	
TX 0	24	36	117	181	378	643	1180	2150	3140	4155	5155	4861	3924	2931	1850	832	354	140	46	15	1	0	
TX	252	432	1244	1992	3858	6678	11978	21638	31208	41148	51014	49176	39226	28978	18274	8414	3614	1424	444	174	14	08	
TY 0	5	6	40	103	248	519	1025	2005	2972	4017	5035	4982	3944	2980	1976	964	463	241	99	58	13	3	
TY	54	110	442	932	2442	4938	9934	19940	29850	39898	50040	50146	39966	29970	19944	9968	4896	2440	948	496	100	48	

N=1024

EMPIRICAL CUMULATIVE PROBABILITY OF THE SAME VALUE OF A STATISTIC WHICH HAS NORMAL-THEORY CUMULATIVE PROBABILITY GIVEN BY COLUMN HEADING (DECIMAL POINTS AND PRECEDING ZEROS OMITTED)

		LEFT-TAIL CUMULATIVE PROBABILITIES											RIGHT-TAIL CUMULATIVE PROBABILITIES											
		0005	0010	0050	0100	0250	0500	1000	2000	3000	4000	5000	5000	4000	3000	2000	1000	0500	0250	0100	0050	0010	0005	CHI-SQ.
ZX	1	3	5	33	79	224	479	1011	1949	2962	4012	5007	5009	3967	2952	2001	1051	557	284	113	51	14	4	33.43
ZY	1	5	14	51	103	249	486	968	1974	2978	3968	4977	5035	4013	3002	2024	1069	539	254	111	55	13	3	21.63
ZX	2	4	7	42	89	216	486	982	2014	3072	4116	5123	4901	3935	2919	1967	991	526	285	128	74	22	8	43.59
ZY	2	4	7	41	89	241	509	1009	2041	3032	4015	4984	5023	4036	2986	1940	939	463	235	95	40	9	5	17.26
ZX	3	1	3	31	82	234	484	991	1994	3043	4028	5056	4959	3926	2936	1960	1012	512	270	114	54	8	4	19.63
ZY	3	1	10	54	95	244	471	964	1951	2960	3958	4907	5098	4041	3019	1964	1006	484	242	80	35	5	2	29.78
ZX	4	1	6	38	93	225	469	978	2018	3030	4112	5147	4860	3908	2888	1965	977	519	272	117	67	17	8	37.58
ZY	4	10	13	41	100	243	496	1004	2036	3062	4005	4985	5023	4028	2999	1995	1016	495	231	95	48	10	4	22.07
ZX	1234	2	10	43	87	227	477	993	2029	3027	4067	5109	4900	3931	2970	2006	1000	515	283	121	58	11	6	22.65
ZY	1234	3	3	40	80	211	463	965	1993	3013	4035	5022	4981	3999	2989	1967	1026	505	257	95	48	6	1	22.12
TX	1	15	28	96	171	345	630	1133	2049	3027	4019	5007	5009	3955	2915	1913	904	415	186	53	26	2	1	112.78
TY	1	8	14	50	105	247	482	968	1971	2994	3959	4977	5035	4014	3013	2030	1060	535	256	109	55	10	3	15.16
TX	2	18	34	108	165	343	644	1140	2115	3123	4126	5123	4901	3915	2890	1868	856	408	190	70	28	2	1	134.48
TY	2	4	9	46	84	231	516	1006	2043	3028	4004	4984	5023	4036	2988	1935	942	465	225	93	42	9	5	21.68
TX	3	16	25	103	170	350	635	1141	2088	3103	4033	5056	4959	3908	2895	1867	862	386	172	49	18	1	1	131.14
TY	3	4	10	54	99	244	467	963	1949	2965	3950	4907	5098	4035	3024	1961	997	490	248	80	31	5	2	26.24
TX	4	19	28	111	172	357	614	1127	2108	3096	4119	5147	4860	3890	2843	1857	833	408	170	63	23	4	4	147.86
TY	4	7	13	40	101	240	497	998	2024	3068	4007	4985	5023	4025	3009	1985	1018	496	231	90	48	9	5	21.53
TX	1234	12	19	70	131	288	548	1064	2066	3063	4075	5109	4900	3922	2948	1955	940	449	234	83	33	3	0	33.33
TY	1234	3	3	42	81	212	468	960	2001	3006	4028	5022	4981	3996	2998	1964	1024	499	257	97	48	6	1	23.69
ZX	0	0	3	32	77	225	472	1017	2017	3047	4100	5078	4934	3863	2856	1915	1003	521	289	125	68	21	11	
ZX		18	48	352	840	2248	4780	9968	19984	30308	40736	50822	49326	39198	29102	19616	10068	5270	2800	1194	648	164	70	
ZY	0	7	8	54	101	239	474	1016	1998	3067	4035	5054	4956	4050	2979	1949	978	500	233	97	60	10	4	
ZY		54	104	482	976	2432	4872	9922	20000	30198	39962	49814	50270	40336	29970	19744	10016	4962	2390	956	476	94	36	
TX	0	13	25	94	163	351	631	1166	2110	3102	4115	5078	4934	3847	2823	1813	855	404	181	64	31	4	2	
TX		162	280	1024	1682	3492	6308	11414	20940	30902	40824	50822	49328	39030	28732	18636	8620	4042	1798	598	252	26	18	
TY	0	6	8	53	98	239	465	1020	1992	3074	4031	5054	4956	4041	2989	1954	971	501	235	99	61	10	5	
TY		58	108	486	974	2402	4894	9910	19968	30258	39902	49814	50270	40302	30046	19730	9976	4974	2390	942	474	86	40	

EMPIRICAL CUMULATIVE PROBABILITY (DECIMAL POINTS AND PRECEDING ZEROS OMITTED) OF Z VALUE

WHICH HAS ONE-TAIL NORMAL-THEORY CUMULATIVE PROBABILITY OF 0001

	LEFT-TAIL CUMULATIVE PROBABILITIES										RIGHT-TAIL CUMULATIVE PROBABILITIES									
N=	2	4	8	16	32	64	128	256	512	1024	2	4	8	16	32	64	128	256	512	1024
ZYY 2NN	1	0	0	0	1	1	1	1	1	2	2	0	1	0	5	0	0	1	1	0
ZYY 3NN	1	0	0	1	2	0	1	2	1	1	2	2	3	0	3	0	0	2	1	0
ZYY 2N2N	3	1	0	0	0	0	2	0	1	1	3	1	0	1	3	1	3	0	2	0
ZXX 2NN	53	21	20	5	2	5	3	2	1	1	2	0	1	0	2	0	0	0	1	1
ZXX 3NN	68	30	23	16	1	6	4	4	1	3	1	0	0	0	0	0	0	1	1	0
ZXX 2N2N	6	5	1	0	1	0	0	1	0	0	14	2	4	0	2	1	2	1	1	1
ZXY 2NN	0	0	0	1	0	0	2	2	1	1	11	10	5	1	4	1	0	3	1	3
ZXY 3NN	0	0	0	2	1	1	2	2	1	0	7	6	3	1	3	1	3	2	1	3
ZXY 2N2N	0	0	0	0	0	0	C	0	0	0	24	17	5	3	5	1	1	2	1	0
ZYX 2NN	61	27	25	7	3	3	5	3	3	3	0	0	0	0	0	0	0	0	0	1
ZYX 3NN	79	35	29	12	2	4	3	4	3	3	0	0	0	0	0	0	0	0	0	0
ZYX 2N2N	24	10	6	3	2	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0
ZXA 2NN	5	3	1	0	0	0	2	0	0	0	25	18	10	3	7	3	5	3	1	3
ZXA 3NN	18	8	5	1	1	3	3	1	0	0	11	8	4	2	4	2	1	4	0	0
ZXA 2N2N	0	0	0	0	0	0	0	0	0	1	27	24	12	4	8	1	3	2	2	2
ZAX 2NN	89	42	34	18	5	6	4	5	2	3	0	0	0	0	0	0	0	0	1	0
ZAX 3NN	93	45	37	22	7	7	4	4	2	3	0	0	0	0	0	0	0	0	1	0
ZAX 2N2N	34	20	12	5	2	0	0	2	1	1	0	0	1	0	0	0	0	0	0	0
ZXB 2NN	0	0	0	0	0	0	2	0	0	1	30	26	9	2	6	1	1	4	1	3
ZXB 3NN	0	0	0	1	1	0	2	2	0	1	11	18	6	5	4	1	1	3	0	4
ZXB 2N2N	0	0	0	0	0	0	0	0	0	1	39	28	15	4	7	5	5	3	3	1
ZBX 2NN	93	45	37	17	5	7	3	6	2	5	0	0	0	0	0	0	0	0	0	0
ZBX 3NN	98	47	38	25	5	7	3	6	3	4	0	0	0	0	0	0	0	0	1	0
ZBX 2N2N	39	23	11	6	3	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
ZYA 2NN	16	9	7	4	1	2	2	1	2	2	0	0	0	0	0	0	1	1	0	1
ZYA 3NN	37	12	14	6	4	2	2	2	3	5	0	0	0	0	0	0	0	0	0	1
ZYA 2N2N	8	5	3	0	2	1	0	0	2	0	1	0	1	0	0	1	3	2	1	1
ZAY 2NN	0	1	0	2	0	0	0	1	1	1	7	1	2	1	4	0	1	2	1	0
ZAY 3NN	1	1	1	3	0	0	0	1	1	0	5	2	2	1	2	0	3	2	1	0
ZAY 2N2N	0	0	0	1	0	0	0	0	1	0	9	3	0	3	3	1	1	1	0	0
ZX NN	19	9	10	1	1	2	1	0	0	1	25	13	2	4	2	0	0	1	3	1
ZY NN	2	0	1	1	0	2	0	2	0	1	0	1	3	1	2	0	1	1	1	3
ZX 1	0	0	0	0	0	0	0	0	0	0	111	44	32	25	8	2	4	6	1	2
ZY 1	0	0	1	0	1	1	1	1	1	1	1	2	3	1	0	0	0	1	1	0
ZX 2	0	0	0	0	0	0	0	0	0	0	108	53	40	20	11	9	2	3	5	2
ZY 2	1	0	1	0	2	0	2	0	1	1	2	1	3	0	0	3	0	1	2	2
ZX 3	0	0	0	0	0	0	0	0	0	0	109	44	30	16	10	7	3	6	0	1
ZY 3	0	0	0	5	1	1	2	1	1	0	0	1	1	0	1	1	1	3	2	0
ZX 4	0	0	0	0	0	0	0	0	1	0	109	49	39	29	10	13	3	6	2	4
ZY 4	2	1	2	0	1	1	1	2	1	1	4	6	1	2	0	1	0	0	1	0
ZX 1234	0	0	0	0	0	0	0	0	0	0	63	24	12	10	7	4	2	5	1	0
ZY 1234	6	5	1	6	0	1	2	0	3	1	6	3	0	3	1	0	0	1	1	0
ZX0	0	0	0	0	0	0	0	0	0	0	115	61	35	25	22	6	3	2	2	3
ZY0	3	1	3	0	1	0	0	0	1	2	1	1	0	1	0	1	1	0	2	0
ZX	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	110.4	50.2	35.2	23.0	12.2	7.4	3.0	4.6	2.0	2.4
ZY	1.2	0.4	1.4	1.0	1.2	0.6	1.2	0.8	1.0	1.0	1.6	1.0	1.6	0.8	0.2	1.2	0.4	1.0	1.6	0.4

NOTES: CUMULATIVE PROBABILITY FOR ANY ENTRY = ENTRY X 10^{-4}

ENTRY FOR ZX = $\frac{1}{5}\sum_{i=0}^{4}$ ENTRY FOR ZXi, AND SIMILARLY FOR ZY

APPENDIX V

Tolerance Limits For Empirical Cumulative Probabilities

The following table gives upper and lower 95%, 99%, and 99.9% tolerance limits for an empirical cumulative probability, ρ, about a true cumulative probability of $\bar{\rho}$. The table therefore gives an indication of the precision with which the _empirical_ significance level, ρ, corresponding to an unspecified normal-theory significance level of α, estimates the _true_ significance level, $\bar{\rho}$, corresponding to the same, unspecified, normal-theory significance level of α. Or, by substituting α for $\bar{\rho}$, the table gives the normal-theory tolerance limits for an empirical significance level, ρ, about a true (i.e., normal-theory) significance level of α.

Almost all of the table entries were obtained by means of the Chi Square approximation to the binomial, using Yates' correction for continuity. Thus, deviations of observed from expected probabilities were obtained by solving for $|k\rho - k\bar{\rho}|$ the formula $\chi_T^2 = \frac{(|k\rho - k\bar{\rho}| - 1/2)^2}{k\bar{\rho}} + \frac{(|k\rho - k\bar{\rho}| - 1/2)^2}{k - k\bar{\rho}}$, where k is the total number of observations comprising the empirical sampling distribution (i.e., either 10,000 or 50,000) and χ_T^2 is the value of χ^2 appropriate to a tolerance level of T. The integral part of $|k\rho - k\bar{\rho}|$ was then solved for ρ, and this value is the table entry.

When $\bar{\rho} = .0001$ and $k = 10{,}000$, the expected frequency, $k\bar{\rho}$, is 1, which is too small for the chi square approximation to be very accurate. Therefore, upper and lower tolerance limits for ρ when $\bar{\rho} = .0001$ and the empirical distribution consists of 10,000 values were obtained directly from the binomial. Let B stand for the point probability of one success in 10,000 trials of a binomial event whose probability of success in a single trial is .0001. Then if r is the number of successes in 10,000 trials, $B = Pr\,(r = 1) = \binom{10{,}000}{1} (.0001)^1 (.9999)^{9999}$. It can be shown that for small values of h, $Pr\,(r = h) \cong \frac{B}{h!}$. Therefore, disregarding the probability that $r \geq 9$ as negligible, the following approximate equation can be solved for B: $1 = \sum_{h=0}^{8} \frac{B}{h!}$. The resulting approximate value of B can be divided by h! to yield point probabilities, and these can be summed to yield cumulative probabilities. In essentially this manner (actually $Pr\,(r = h)$ was expressed in terms of B and $1 = \sum_{h=0}^{8} Pr(r)$ was solved for B), the following table was obtained for the probabilities of $\rho \geqq \frac{h}{10,000}$ in a sampling distribution of 10,000 observations when the true value of ρ, i.e., $\bar{\rho}$, is .0001.

h	$Pr\ (r \geqq h) = Pr\ (\rho \geqq \frac{h}{10,000})$
0	1.00000
1	.63214
2	.26424
3	.08029
4	.01898
5	.00366
6	.00059
7	.00008
8	.00001

This table was used to provide the entries corresponding to $\bar{\rho} = .0001$ and a distribution of 10,000 values in Table VIII. These entries are probably much more accurate than are their near neighbors, since none of the tolerance probability is wasted in representing (with nonzero probabilities) the impossible events of ρ's less than zero.

Table VIII

UPPER (U) AND LOWER (L) TOLERANCE LIMITS FOR ρ ABOUT A TRUE VALUE OF $\bar{\rho}$

$\bar{\rho}$		Distribution Based upon 10,000 Empirical Values			Distribution Based upon 50,000 Empirical Values		
		95%	99%	99.9%	95%	99%	99.9%
.0001	U	.0003	.0004	.0005	.00018	.00022	.00024
	L	.0000	.0000	.0000	.00002	.00000	.00000
.0005	U	.0009	.0011	.0012	.00070	.00076	.00082
	L	.0001	.0000	.0000	.00030	.00024	.00018
.001	U	.0016	.0018	.0020	.00128	.00136	.00146
	L	.0004	.0002	.0000	.00072	.00064	.00054
.005	U	.0064	.0068	.0073	.00562	.00582	.00604
	L	.0036	.0032	.0027	.00438	.00418	.00396
.01	U	.0120	.0126	.0133	.01088	.01114	.01146
	L	.0080	.0074	.0067	.00912	.00886	.00854
.025	U	.0281	.0290	.0301	.02636	.02680	.02730
	L	.0219	.0210	.0199	.02364	.02320	.02270
.05	U	.0543	.0556	.0572	.05192	.05252	.05320
	L	.0457	.0444	.0428	.04808	.04748	.04680
.10	U	.1059	.1077	.1099	.10262	.10346	.10442
	L	.0941	.0923	.0901	.09738	.09654	.09558
.20	U	.2078	.2103	.2132	.20350	.20460	.20588
	L	.1922	.1897	.1868	.19650	.19540	.19412
.30	U	.3090	.3118	.3151	.30402	.30528	.30674
	L	.2910	.2882	.2849	.29598	.29472	.29326
.40	U	.4096	.4126	.4161	.40430	.40564	.40720
	L	.3904	.3874	.3839	.39570	.39436	.39280
.50	U	.5098	.5129	.5165	.50438	.50576	.50736
	L	.4902	.4871	.4835	.49562	.49424	.49264

APPENDIX VI

Influence of Population Moments upon the Robustness of the Sample Mean

In the body of this report, robustness data were presented for 23 Z statistics of the form

$$ZPQ\ C_1NC_2N = \frac{\bar{P} - \bar{Q}}{\sqrt{\frac{\sigma_P^2}{C_1N} + \frac{\sigma_Q^2}{C_2N}}} = \frac{\frac{\sum_{i=1}^{C_1N} P_i}{C_1N} - \frac{\sum_{j=1}^{C_2N} Q_j}{C_2N}}{\sqrt{\frac{\sigma_P^2}{C_1N} + \frac{\sigma_Q^2}{C_2N}}} .$$

Since the Z statistic's denominator is a constant for a given value of N, the Z statistic is simply a linear transformation upon its own numerator and the robustness of Z is equally the robustness of its numerator. On pages 96 and 97 (see also page 166) it was shown, in effect, that the numerator of ZPQ C_1NC_2N can be legitimately regarded as the mean of N values of the variable $\bar{P}_{C_1} - \bar{Q}_{C_2}$, where $\bar{P}_{C_1} = \frac{\sum_{i=1}^{C_1} P_i}{C_1}$ and $\bar{Q}_{C_2} = \frac{\sum_{j=1}^{C_2} Q_j}{C_2}$. So robustness data on ZPQ C_1NC_2N are equally data on the robustness of the mean of N observations upon the single variable $\bar{P}_{C_1} - \bar{Q}_{C_2}$. These robustness data therefore show the Central Limit effect upon the mean of a sample of size N drawn from the $\bar{P}_{C_1} - \bar{Q}_{C_2}$ population.

The central moments of the $\bar{P}_{C_1} - \bar{Q}_{C_2}$ populations can be easily derived using only algebra and the rules for expected values. The r th central moment of the $\bar{P}_{C_1} - \bar{Q}_{C_2}$ population is

$$\mu_r(\bar{P}_{C_1} - \bar{Q}_{C_2}) = E\left[\bar{P}_{C_1} - \bar{Q}_{C_2} - E(\bar{P}_{C_1} - \bar{Q}_{C_2})\right]^r$$

$$= E\left[(\bar{P}_{C_1} - \mu_P) - (\bar{Q}_{C_2} - \mu_Q)\right]^r$$

$$= E\Big[(\bar{P}_{C_1} - \mu_P)^r - \binom{r}{r-1}(\bar{P}_{C_1} - \mu_P)^{r-1}(\bar{Q}_{C_2} - \mu_Q)$$

$$+\binom{r}{r-2}(\bar{P}_{C_1} - \mu_P)^{r-2}(\bar{Q}_{C_2} - \mu_Q)^2 + \ldots\ldots$$

$$+ (-1)^i\binom{r}{r-i}(\bar{P}_{C_1} - \mu_P)^{r-i}(\bar{Q}_{C_2} - \mu_Q)^i + \ldots + (-1)^r(\bar{Q}_{C_2} - \mu_Q)^r\Big]$$

$$= \mu_r(\bar{P}_{C_1}) - \binom{r}{r-1} \mu_{r-1}(\bar{P}_{C_1}) \mu_1(\bar{Q}_{C_2}) + \binom{r}{r-2} \mu_{r-2}(\bar{P}_{C_1}) \mu_2(\bar{Q}_{C_2})$$

$$+ \dots + (-1)^i \binom{r}{r-i} \mu_{r-i}(\bar{P}_{C_1}) \mu_i(\bar{Q}_{C_2}) + \dots + (-1)^r \mu_r(\bar{Q}_{C_2})$$

where the $\binom{r}{r-i}$ are the binomial coefficients, $\frac{r!}{(r-i)!\, i!}$. This equation can be solved if the central moments of $\bar{P}_{C_1}$ and $\bar{Q}_{C_2}$ are known, i.e., knowing the r th central moment of the mean of N values of a variable X, where N stands for C_1 or C_2 and X stands for P or Q respectively. These central moments can be derived as follows.

$$\mu_r(\bar{X}) = E\,[\bar{X} - E(\bar{X})]^r = E\left[\frac{\sum_{i=1}^{N} X_i}{N} - \mu\right]$$

$$= E\left[\frac{\sum_{i=1}^{N} X_i - N\mu}{N}\right]^r = E\left[\frac{\sum_{i=1}^{N} (X_i - \mu)}{N}\right]^r$$

$$= \frac{1}{N^r}\; E\,[Y_1 + Y_2 + \dots + Y_i + \dots + Y_N]^r$$

where $Y_i = X_i - \mu$. Now the general term in the expansion of $[Y_1 + Y_2 + \dots + Y_i + \dots + Y_N]^r$ is $\left[C_1, C_2, \overset{r}{\dots}, C_i, \dots C_N\right] Y_1^{C_1} Y_2^{C_2} \dots Y_i^{C_i} \dots Y_N^{C_N}$ where $\left[C_1, C_2, \overset{r}{\dots}, C_i, \dots C_N\right]$ is the multinomial coefficient $\frac{r!}{C_1!\, C_2! \dots C_i! \dots C_N!}$ in which $\sum_{i=1}^{N} C_i = r$ so that no more than r of the C_i's are nonzero values. Let n be the number of C_i's which are not zeros. The expected value of the general term will be $\left[C_1, C_2, \overset{r}{\dots}, C_i, \dots, C_N\right] \mu_{C_1}(X)\, \mu_{C_2}(X) \dots \mu_{C_i}(X) \dots \mu_{C_N}(X)$ which is independent of the particular set of n subscripts associated with Y's having nonzero exponents. Thus, irrespective of the subscripts involved, all terms having exactly the same set of n nonzero exponents will have the same expected value. Let P be the number of distinguishable permutations of exponents among the set of n Y's having nonzero exponents. Each such distinguishable permutation will result in a <u>different</u> term having the <u>same</u> set of subscripts associated with nonzero exponents and having the <u>same</u> expected value. And there are $\binom{N}{n}$ different ways in which different sets of n subscripts can be chosen from among the N subscripts to represent the set of n Y's having nonzero exponents, and for each such "way" there is a different term having the same expected value. Therefore for a given set of n nonzero exponents, C_i's, there will be $P\binom{N}{n}$ terms of the form $\left[C_1, C_2, \ldots^{r}, C_i, \ldots, C_N\right] Y_1^{C_1} Y_2^{C_2} \ldots Y_i^{C_i} \ldots Y_N^{C_N}$ which have the same expected value, $\left[C_1, C_2, \ldots^{r}, C_i, \ldots C_N\right] \mu_{C_1}(X)\, \mu_{C_2}(X) \ldots \mu_{C_i}(X) \ldots \mu_{C_N}(X)$.

Therefore the general term of $\mu_r(\bar{X}) = \frac{1}{N^r} E\left[Y_1 + Y_2 + \ldots + Y_i + \ldots + Y_N\right]^r$ is

$$\frac{1}{N^r}\left\{P\binom{N}{n}\left[C_1, C_2, \ldots^{r}, C_i, \ldots C_N\right] \mu_{C_1}(X)\, \mu_{C_2}(X) \ldots \mu_{C_i}(X) \ldots \mu_{C_N}(X)\right\}$$

and there will be as many such terms as there are different combinations of nonzero integers, (i.e., nonzero C_i's), whose sum is r. Thus, for example,

$$\mu_4(\bar{X}) = \frac{1}{N^4}\left\{1\binom{N}{1}\left[{4 \atop 4}\right]\mu_4(X) + 2\binom{N}{2}\left[{4 \atop 3,1}\right]\mu_3(X)\,\mu_1(X) + 1\binom{N}{2}\left[{4 \atop 2,2}\right]\mu_2(X)\mu_2(X)\right.$$

$$\left. + 3\binom{N}{3}\left[{4 \atop 2,1,1}\right]\mu_2(X)\mu_1(X)\mu_1(X) + 1\binom{N}{4}\left[{4 \atop 1,1,1,1}\right]\mu_1(X)\mu_1(X)\mu_1(X)\mu_1(X)\right\}$$

and, since $\mu_1(X) = E\ (X - \mu) = 0$,

$$\mu_4(\bar{X}) = \frac{N\mu_4(X) + 0 + 3N(N-1)\ \mu_2^{\,2}(X) + 0 + 0}{N^4}$$

$$= \frac{\mu_4(X) + 3(N-1)\ \sigma^4(X)}{N^3}$$

All the required $\mu_r(\bar{P}_{C_1})$ and $\mu_r(\bar{Q}_{C_2})$ were derived in this way. In connection with a different problem, the true distribution of each of the $\bar{P}_{C_1}$ and $\bar{Q}_{C_2}$ populations had been obtained by the machine and from these distributions the first ten central moments had been calculated directly by the machine. These served as a check for the central moments obtained by "hand" calculation from the derived equations.

The central moments for different populations can be made comparable by standardizing them. This consists of dividing each μ_r by $\mu_2^{r/2}$, i.e., by the population standard deviation raised to the r th power. This produces the same effect as would have been accomplished had the central moments been calculated on standardized population variates rather than upon the raw variate values. Let X be the raw variate value and $Z = \frac{X - \mu}{\sigma}$ the standardized variate value. Then

$$\mu_r(X) = E\,(X-\mu)^r$$

$$\mu_r(Z) = E\,[Z-E(Z)]^r = E\,[Z-0]^r = E\,[Z^r]$$

$$= E\left[\frac{X-\mu}{\sigma}\right]^r = \frac{E\,(X-\mu)^r}{\sigma^r} = \frac{\mu_r(X)}{\sigma^r} = \frac{\mu_r(X)}{\mu_2^{r/2}(X)} = \alpha_r(X),$$ in common

notation for standardized central moments.

Standardizing the central moment obtained in the previous example we have

$$\alpha_4(\bar{X}) = \frac{\mu_4(\bar{X})}{\sigma^4(\bar{X})} = \frac{\mu_4(\bar{X})}{\frac{\sigma^4(X)}{N^2}} = \frac{\mu_4(X) + 3\,(N-1)\,\sigma^4(X)}{\frac{N^3\,\sigma^4(X)}{N^2}} = \frac{\alpha_4(X) - 3}{N} + 3$$

Once the central moments of a population have been standardized, they can be meaningfully compared with the standardized central moments of a normal population. Let ψ_r be the r th standardized central moment, $\alpha_r = \frac{\mu_r}{\sigma^r}$, of a normal population. (When r is odd, $\psi_r = 0$; when r is even, ψ_r is the product (1) (3) (5) ... (r-1) of all the odd integers smaller than r.) And let

$$\partial_r(X) = \alpha_r(X) - \psi_r = \frac{\mu_r(X)}{\sigma^r(X)}$$ minus the value of $\frac{\mu_r}{\sigma^r}$ for a normal population.

Then $\partial_r(X)$ indicates the deviation or departure of the X population's r th standardized central moment from the corresponding standardized central moment of a normally distributed population, and ∂_r therefore acts as an index of departure from its normal-theory value by the r th standardized central moment. Continuing the previous example,

$$\partial_4(\bar{X}) = \alpha_4(\bar{X}) - \psi_4 = \alpha_4(\bar{X}) - 3 = \frac{\alpha_4(X) - 3}{N} + 3 - 3 = \frac{\alpha_4(X) - 3}{N} = \frac{\partial_4(X)}{N}.$$

If, for a given population, its mirror-image (about an axis through its mean) is substituted, the central moments, μ_r's, of the substituted population will be the same as those of the original population except that the algebraic sign of the odd central moments will be reversed. And the distribution of the sample mean for the substitute population will be the mirror-image of that for the original population. Therefore we have moments, and data on the robustness of the sample mean, not only for the twenty-three $\bar{P}_{C_1} - \bar{Q}_{C_2}$ populations but also, in effect, for their mirror-images. However, since one member of a pair of mirror-image populations (and the data on the robustness of its sample mean) provides all of the necessary "information" required of the other member (and its sample mean), it would be a redundancy to present both sets of information. It is desirable, however, to use that member of each pair whose third central moment is positive, since this makes both the populations and the robustness data on their sample means more concisely and meaningfully comparable. This was accomplished as follows. For populations, $\bar{P}_{C_1} - \bar{Q}_{C_2}$, whose third central moment was negative, (a) the algebraic signs of all odd central moments were reversed, (b) robustness data for the left tail of the sample mean was redesignated "right tail" and vice versa, (c) the population was redesignated $\bar{Q}_{C_2} - \bar{P}_{C_1}$. In addition to the twenty-three $\bar{P}_{C_1} - \bar{Q}_{C_2}$ type populations, a twenty-fourth population was provided by the X population for which moments can also be calculated and for whose sample mean robustness data is also available. In order of increasing values of the third standardized central moment, the twenty-four populations are: $\bar{X}_2 - \bar{X}_2$, $\bar{A}_3 - Y$, $\bar{A}_2 - Y$, $A - \bar{X}_3$, $\bar{A}_2 - \bar{Y}_2$, $\bar{X}_3 - Y$, $\bar{X}_2 - Y$, $A - \bar{Y}_2$, $\bar{X}_2 - A$, $\bar{X}_3 - B$, $\bar{X}_2 - \bar{Y}_2$, $A - \bar{Y}_3$, $\bar{X}_2 - B$, $X - \bar{X}_2$, $\bar{X}_2 - \bar{A}_2$, $\bar{X}_2 - \bar{B}_2$, $X - \bar{Y}_2$, $X - \bar{X}_3$, $X - \bar{Y}_3$, $X - \bar{A}_2$, $X - \bar{B}_2$, $X - \bar{A}_3$, $X - \bar{B}_3$, X. The eighth and ninth populations in the above order had identical third (and also identical fifth) standardized central moments. (And it should be noted that the five populations of the type $\bar{P}_2 - \bar{Q}_2$ are not the same as the five P - Q populations dealt with on pages 80-90.) For each of these twenty-four populations the first seven central moments were calculated, and converted to ∂_i's, and data already presented in figure 3 and in the upper portion of figures 16-38 provided the required robustness data in the form of the ratios between ρ and α for the mean of N observations drawn, in effect, from one of the twenty-four $\bar{P}_{C_1} - \bar{Q}_{C_2}$ or X populations.

Figure 71 plots the ratio between ρ and α for the mean of a sample of $N = 2$ observations, as ordinates, against the standardized third central moment, $\alpha_3 = \partial_3$, of the sampled population, as abscissa. And below this is plotted the difference between actual and normal-theory values of the fourth to seventh standardized central moments, i.e., the ∂_i, as ordinates, against the difference between the actual and normal-theory values of the third standardized central moment as abscissa, i.e., against the same abscissa scale as used in the top portion. Figure 72 does the same thing except that $N = 16$, and figure 73 is entirely analogous except that in place of ρ at a single value of N, it uses the average value of ρ taken over the ten N values, 2, 4, 8, 16, 32, 64, 128, 256, 512, and 1024. The lower portions of the three figures (showing the relationship between the various ∂_i and ∂_3) are identical.

All three figures show a tendency for nonrobustness to increase (i.e., for ρ or $\bar{\rho}$ to depart increasingly from α or, equivalently, for the ratio between ρ and α to depart increasingly from 1.00) as the third standardized central moment increases above the normal-theory value of zero. Departures of the ρ/α and α/ρ curves from monotonicity-of-increase with increasing values of ∂_3 are associated with departures of the higher order ∂_i from monotonicity-of-increase with increasing ∂_3. And the extent of the former departures from monotonicity tends to diminish as N increases. That this should be the case can be seen from the following equations which give the ∂_i for the sample mean, $\bar{X}$, in terms of the ∂_i for the sampled population (and which were derived in the manner indicated earlier).

$$\partial_3(\bar{X}) = \frac{\partial_3(X)}{\sqrt{N}}$$

$$\partial_4(\bar{X}) = \frac{\partial_4(X)}{N}$$

$$\partial_5(\bar{X}) = \frac{\partial_5(X) + 10\ (N-1)\ \partial_3(X)}{N\sqrt{N}}$$

$$\partial_6(\bar{X}) = \frac{\partial_6(X) + 5\ (N-1)\left[3\ \partial_4(X) + 2\ \partial_3{}^2(X)\right]}{N^2}$$

RELATIONSHIP BETWEEN THE ROBUSTNESS OF THE SAMPLE MEAN AT N = 2 AND THE EXTENT TO WHICH THE STANDARDIZED MOMENTS OF THE SAMPLED POPULATION DIFFER FROM THOSE OF A NORMAL DISTRIBUTION (FOR 24 POPULATIONS WHOSE FIRST SEVEN STANDARDIZED CENTRAL MOMENTS EQUAL OR EXCEED THOSE OF A NORMAL DISTRIBUTION)

ρ = ONE-TAILED CUMULATIVE PROBABILITY (IN AN EMPIRICAL SAMPLING DISTRIBUTION CONTAINING AT LEAST 10,000 VALUES OF THE SAMPLE MEAN) OF THAT VALUE OF THE SAMPLE MEAN WHOSE NORMAL-THEORY ONE-TAILED CUMULATIVE PROBABILITY IS α

$\delta_i = \frac{\mu_i}{\sigma^i}$ MINUS NORMAL-THEORY VALUE OF $\frac{\mu_i}{\sigma^i}$, WHERE μ_i IS THE ith CENTRAL MOMENT AND σ IS THE STANDARD DEVIATION OF THE SAMPLED POPULATION

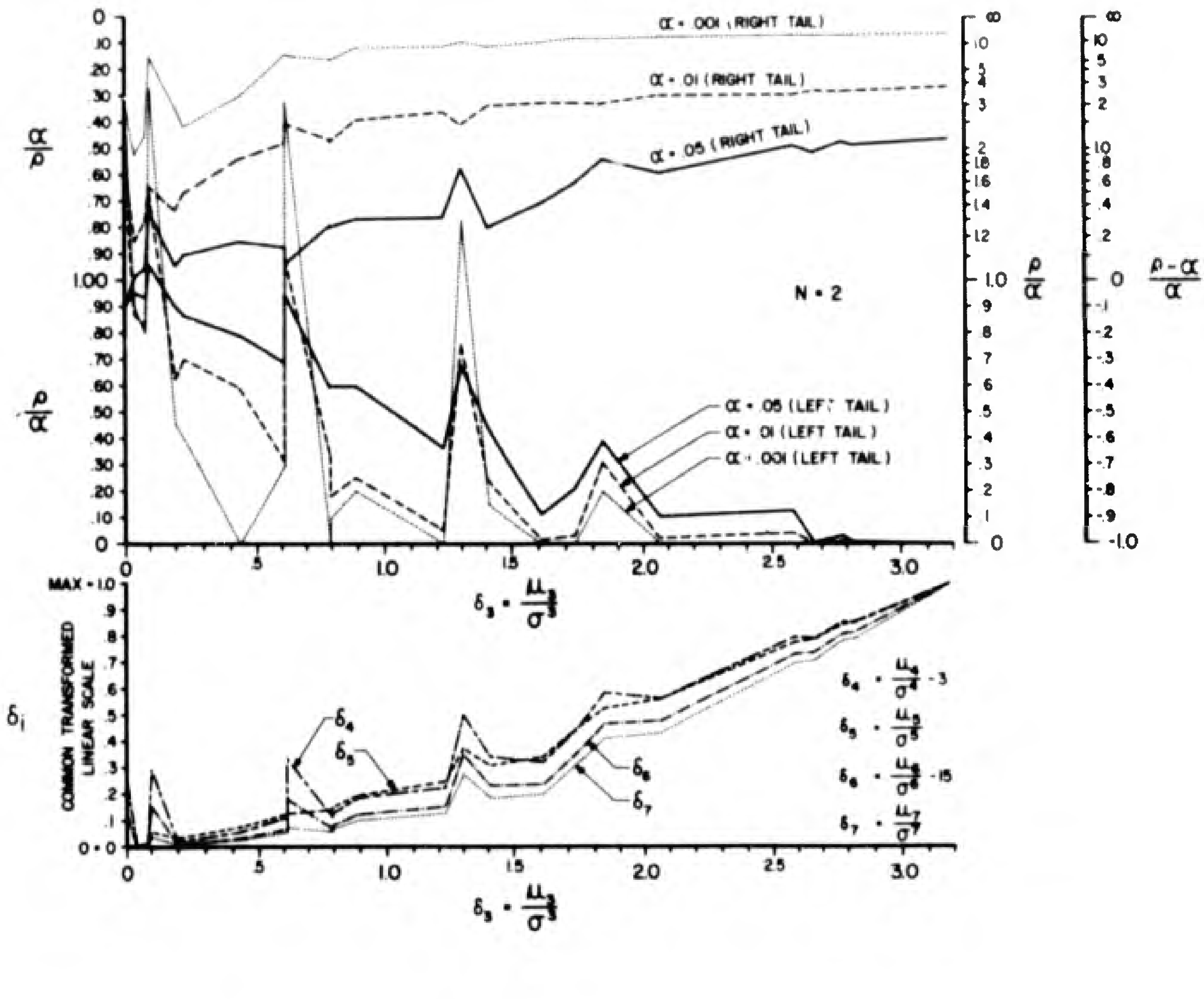

Figure 71. Relationship between Population Moments (especially the third) and the Robustness of the Sample Mean at N = 2

RELATIONSHIP BETWEEN THE ROBUSTNESS OF THE SAMPLE MEAN AT N = 16 AND THE EXTENT TO WHICH THE STANDARDIZED MOMENTS OF THE SAMPLED POPULATION DIFFER FROM THOSE OF A NORMAL DISTRIBUTION (FOR 24 POPULATIONS WHOSE FIRST SEVEN STANDARDIZED CENTRAL MOMENTS EQUAL OR EXCEED THOSE OF A NORMAL DISTRIBUTION)

ρ = ONE-TAILED CUMULATIVE PROBABILITY (IN AN EMPIRICAL SAMPLING DISTRIBUTION CONTAINING AT LEAST 10,000 VALUES OF THE SAMPLE MEAN) OF THAT VALUE OF THE SAMPLE MEAN WHOSE NORMAL-THEORY ONE-TAILED CUMULATIVE PROBABILITY IS α

$\delta_i = \frac{\mu_i}{\sigma^i}$ MINUS NORMAL-THEORY VALUE OF $\frac{\mu_i}{\sigma^i}$, WHERE μ_i IS THE iTH CENTRAL MOMENT AND σ IS THE STANDARD DEVIATION OF THE SAMPLED POPULATION

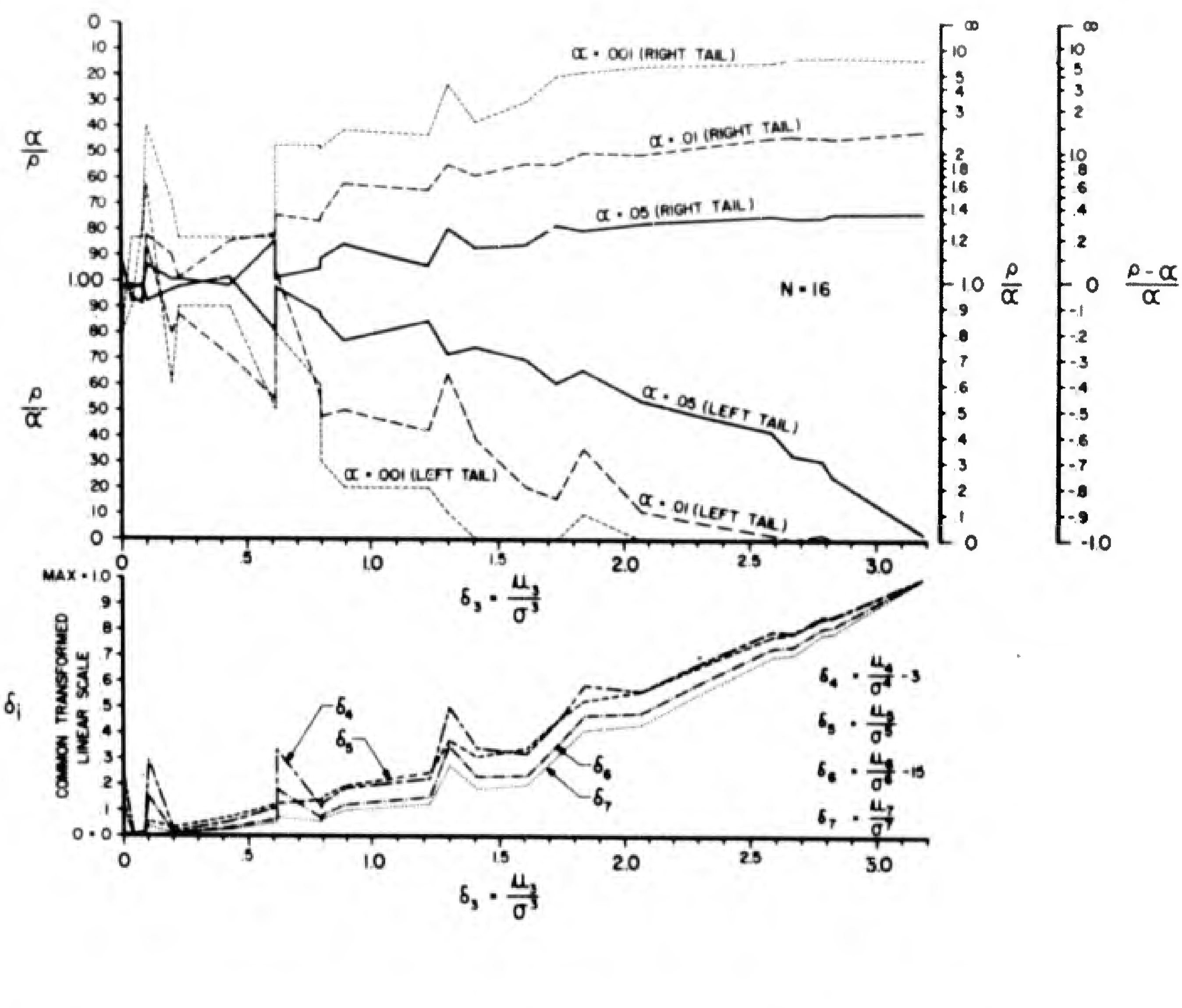

Figure 72. Relationship between Population Moments (especially the third) and the Robustness of the Sample Mean at N = 16

RELATIONSHIP BETWEEN THE ROBUSTNESS OF THE SAMPLE MEAN AND THE EXTENT TO WHICH THE STANDARDIZED MOMENTS OF THE SAMPLED POPULATION DIFFER FROM THOSE OF A NORMAL DISTRIBUTION (FOR 24 POPULATIONS WHOSE FIRST SEVEN STANDARDIZED CENTRAL MOMENTS EQUAL OR EXCEED THOSE OF A NORMAL DISTRIBUTION)

$\bar{\rho}$ = AVERAGE VALUE OF ρ_N, WHERE ρ_N IS THE ONE-TAILED CUMULATIVE PROBABILITY (IN AN EMPIRICAL SAMPLING DISTRIBUTION CONTAINING AT LEAST 10,000 VALUES OF THE SAMPLE MEAN OF N OBSERVATIONS) OF THAT VALUE OF THE SAMPLE MEAN WHOSE NORMAL-THEORY ONE-TAILED CUMULATIVE PROBABILITY IS α, AND THE N VALUES UPON WHICH THE AVERAGE IS BASED ARE 2, 4, 8, 16, 32, 64, 128, 256, 512, AND 1024

$\delta_i = \frac{\mu_i}{\sigma^i}$ MINUS NORMAL-THEORY VALUE OF $\frac{\mu_i}{\sigma^i}$, WHERE μ_i IS THE ITH CENTRAL MOMENT AND σ IS THE STANDARD DEVIATION OF THE SAMPLED POPULATION

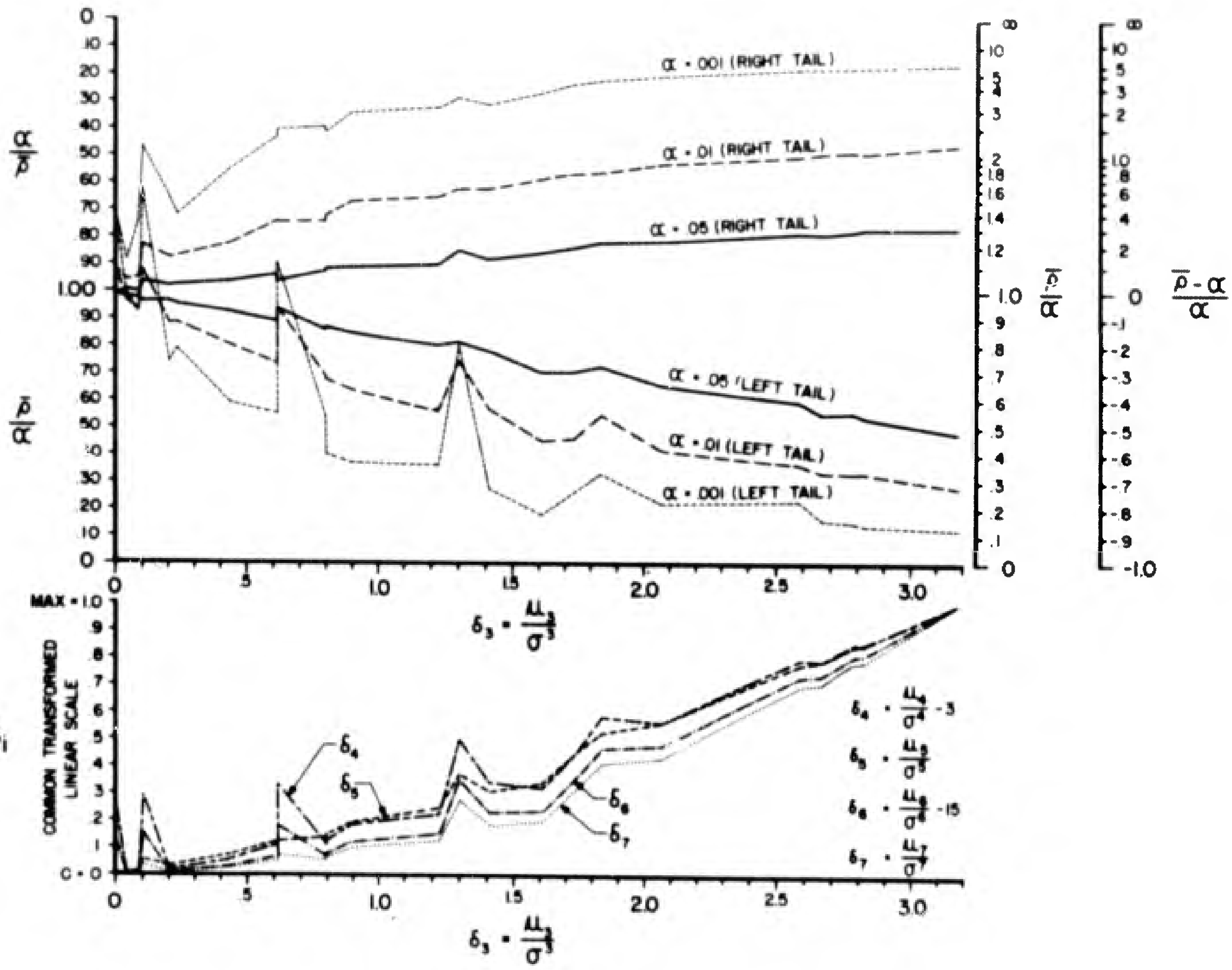

Figure 73. Relationship between Population Moments (especially the third) and the Robustness of the Sample Mean (Averaged over Ten Values of N)

$$\partial_7(\bar{X}) = \frac{\partial_7(X) + 7\ (N-1)\left[\ 3\ \partial_5(X) + 5\ \partial_4(X)\ \partial_3(X) + 15\ (N-1)\ \partial_3(X)\right]}{N^2\ \sqrt{N}}$$

$$\partial_8(\bar{X}) = \frac{1}{N^3}\left\{\partial_8(X) + 7\ (N-1)\left[4\ \partial_6(X) + 8\ \partial_5(X)\ \partial_3(X) + 5\ \partial_4^{\ 2}(X)\right.\right.$$

$$\left.\left. + 30\ (N-1)\ \partial_4(X) + 40\ (N-2)\ \partial_3^{\ 2}(X)\ \right]\ \right\}$$

$$\partial_9(\bar{X}) = \frac{1}{N^3\sqrt{N}}\left\{\partial_9(X) + 2\ (N-1)\left[18\ \partial_7(X) + 42\ \partial_6(X)\ \partial_3(X) + 63\ \partial_5(X)\ \partial_4(X)\right.\right.$$

$$+ 189\ (N-1)\ \partial_5(X) + 630\ (N-2)\ \partial_4(X)\ \partial_3(X) + 140\ (N-2)\ \partial_3^{\ 3}(X)$$

$$\left.\left. + 630\ (N-1)^2\ \partial_3(X)\right]\ \right\}$$

When $r \leqq 4$, $\partial_r(\bar{X})$ depends upon $\partial_r(X)$ and upon no other $\partial_i(X)$. When $r \geqq 5$, all of the $\partial_i(X)$ in the numerator except $\partial_r(X)$ are collectively weighted by N-1 in contributing to $\partial_r(\bar{X})$. (They may also receive further "N-weights" on an individual basis.) And when $r > 6$, additional "N-weights" tend to be given to the lower order ∂_i appearing in the numerator. In fact in deriving the $\partial_r(\bar{X})$ it becomes clear that (especially when r is large) the amount of "N-weight" received by a $\partial_i(X)$ in contributing to $\partial_r(\bar{X})$ tends to increase as i decreases from r-2 to 3. (The function of N constituting the denominator is ignored here, since it does not affect the $\partial_i(X)$ differentially.)

It is clear from the above, therefore, that the larger the value of N, the greater the weight given to the lower-order, relative to that given to the higher-order, $\partial_i(X)$ in contributing to $\partial_r(\bar{X})$, provided that $r \geqq 5$. Thus if sample size is moderately large, the higher-order moments of the distribution of the sample mean, as well as the lower-order moments, will be largely determined by the lower-order moments of the sampled population. (Another factor which tends to produce the same effect, and therefore enhances the effect named, is the tendency for all odd moments of the sampled population to be well correlated with one another, especially with their near neighbors, and likewise for all even moments, so that the multiple correlation between $\partial_r(\bar{X})$ and all $\partial_i(X)$ can often be fairly well represented by that between $\partial_r(\bar{X})$ and only $\partial_3(X)$ and $\partial_4(X)$. Furthermore, if the sampled population is asymmetrical, having a long tail on one side of the mean and a short one on the other, the extreme values on the long tail will tend to account for a large proportion of the magnitude of the higher moments ($r \geq 3$), and consequently the correlation between odd and even moments will tend to be high.) So, at fairly modest values of N, all moments of the distribution of the sample mean, and therefore the shape of the distribution of $\bar{X}$, and therefore the robustness of $\bar{X}$ (and therefore the Central Limit effect upon $\bar{X}$) tend to be largely determined by the third and fourth moments of the sampled population (or, equivalently, by $\partial_3(X)$ and $\partial_4(X)$), especially by the third.

APPENDIX VII

Derivational Note

This appendix serves as a mathematical footnote to statements and derivations appearing on pages 112, 114, 120, 124, 139, 162, and 192, all of which pertain to the substitution of E(d) for d in the ratio $F = \frac{n}{d}$ when the relative variance of d becomes zero.

_If_ E(d) = E(n), _or_ if both E(d) and E(n) are indenendent of absolute sample size, then the fact that $\frac{Var\ (d)}{Var\ (n)} \rightarrow 0$ as $N_{min} \rightarrow \infty$ implies that $\frac{n}{d} \rightarrow \frac{n}{E(d)}$ as $N_{min} \rightarrow \infty$ so that at $N_{min} = \infty$, $F = \frac{n}{E(d)}$. It does _not_ imply it, however, if as $N_{min} \rightarrow \infty$ $\frac{E(d)}{E(n)} \rightarrow 0$ at a rate as great or greater than that at which $\frac{Var\ (d)}{Var\ (n)} \rightarrow 0$.

The sufficient condition, E(d) = E(n), is met: (a) under normal theory (see page 111), (b) when only the normality assumption is violated (see page 121), or (c) when sample sizes all are equal, irrespective of nonnormality or heterogeneity (see page 121). It is not met when sample sizes and population variances both are unequal. In that case, taking the formulas for E(n) and E(d) from page 121, letting N be the smallest of the N_i and r_i be the ratio N_i/N so that $N_i = r_iN$ (i.e., the product of relative, r_i, and minimum absolute, N, sample size) and continuing the derivation of expected values,

$$E(n) = \frac{T \sum_1^c K_i \sigma_i^2 - \sum_1^c K_i N_i \sigma_i^2}{T\ (K-1)}$$

$$= \frac{\left[\sum_1^c K_i\ (r_iN)\right] \sum_1^c K_i \sigma_i^2 - \sum_1^c K_i \cdot (r_iN)\ \sigma_i^2}{\left[\sum_1^c K_i\ (r_iN)\right] (K-1)}$$

$$= \frac{\left[\sum_1^c K_i r_i\right] \sum_1^c K_i \sigma_i^2 - \sum_1^c K_i r_i \sigma_i^2}{(K-1) \sum_1^c K_i r_i}$$

which is independent of N, i.e., of absolute sample size, and is a constant for given K_i, r_i, and σ_i^2, i.e., is constant for constant sampling conditions other than absolute sample size.

$$E(d) = \frac{\sum_1^c K_i (N_i - 1) \sigma_i^2}{\sum_1^c K_i (N_i - 1)} = \frac{N \sum_1^c K_i r_i \sigma_i^2 - \sum_1^c K_i \sigma_i^2}{N \sum_1^c K_i r_i - \sum_1^c K_i}$$

Since all $r_i \geq 1$, it is the case both in the numerator and in the denominator that the subtracted term is no larger than the term by which N is multiplied. Thus the subtracted terms have a relatively small effect upon E(d) even at moderate values of N. But to the extent that the subtracted terms are negligible, the N's in numerator and denominator cancel. And, of course, when $N = \infty$

$$E(d) = \frac{\sum_1^c K_i r_i \sigma_i^2}{\sum_1^c K_i r_i}$$

which, like E(n), is a constant for constant sampling conditions other than absolute sample size. Thus E(d) is a weak function of N at moderate N values, becoming weaker as N increases, and finally reaching complete independence of N when the latter assumes infinitely large values.

As $N \to \infty$, therefore, $\frac{E(d)}{E(n)}$ does not change at as fast a rate as that with which $\frac{Var\ (d)}{Var\ (n)} \to 0$, (the latter rate being $\frac{\text{a constant}}{N}$)and, instead of approaching zero, approaches a substantial (i.e., noninfinitessimal) nonzero constant value. It follows, therefore, that as $N_{min} \to \infty$, $\frac{n}{d} \to \frac{n}{E(d)}$, despite the fact that $E(d) \neq E(n)$.

In the case of the t test, the expected value of the denominator does approach zero as $N \to \infty$. However the expected value of the numerator is already zero at finite sample sizes. Ambiguities can be resolved and derivations simplified by dealing with t^2 rather than t, as done in pages 137-141; since if the denominator of t^2 behaves as a constant in contributing to the mean and variance of t^2, so will the denominator of t in contributing to the mean and variance of t. This has already been done for the two sample t test, in the body of this report, and all of the derivations and conclusions presented in this appendix apply to this test since it is a special case of the F test.

For the one sample t statistic,

$$t^2 = \frac{(\bar{X}-\mu)^2}{\hat{\sigma}^2/N} = \frac{n}{d}$$

$$E(n) = E(\bar{X}-\mu)^2 = \sigma_{\bar{X}}^2 = \sigma^2/N$$

$$E(d) = E(\hat{\sigma}^2/N) = \sigma^2/N = E(n)$$

$$\mathrm{Var}(n) = E\left[(\bar{X}-\mu)^2 - E(\bar{X}-\mu)^2\right]^2 = E\left[(\bar{X}-\mu)^2 - \sigma_{\bar{X}}^2\right]^2$$

$$= E\left[(\bar{X}-\mu)^4 - 2\,\sigma_{\bar{X}}^2\,(\bar{X}-\mu)^2 + \sigma_{\bar{X}}^4\right]$$

$$= \mu_4(\bar{X}) - 2\,\sigma_{\bar{X}}^2\,\sigma_{\bar{X}}^2 + \sigma_{\bar{X}}^4 = \mu_4(\bar{X}) - \sigma_{\bar{X}}^4$$

Substituting for $\mu_4(\bar{X})$, the fourth central moment of the sample mean, its equivalent value in terms of the moments of the sampled population (given on page 278) and doing likewise for $\sigma_{\bar{X}}^4$,

$$\mathrm{Var}(n) = \frac{\mu_4 + 3(N-1)\,\sigma^4}{N^3} - \frac{\sigma^4}{N^2} = \frac{\mu_4 + (2N-3)\,\sigma^4}{N^3}$$

$$\mathrm{Var}(d) = \mathrm{Var}\left(\frac{\hat{\sigma}^2}{N}\right) = \frac{1}{N^2}\,\mathrm{Var}\,(\hat{\sigma}^2)$$

Substituting the formula for Var $(\hat{\sigma}^2)$ given on page 116, (where the symbol ϕ is used for $\mu_4(X)$)

$$\mathrm{Var}(d) = \frac{1}{N^2}\left[\frac{\mu_4 - \sigma^4}{N} + \frac{2\sigma^4}{N(N-1)}\right]$$

$$= \frac{\mu_4 - \left(\frac{N-3}{N-1}\right)\sigma^4}{N^3}$$

$$\frac{\mathrm{Var(d)}}{\mathrm{Var(n)}} = \frac{\mu_4 - \left(\frac{N-3}{N-1}\right)\sigma^4}{\mu_4 + (2N-3)\,\sigma^4}$$

which as $N \to \infty$ becomes

$$\frac{Var(d)}{Var(n)} = \frac{\mu_4 - \sigma^4}{\mu_4 \quad (2^{\infty} - 3)\sigma^4} = \frac{\text{a constant}}{\text{infinity}} = 0$$

Therefore, as $N \to \infty$, $\frac{n}{d} \to \frac{n}{E(d)}$, $t^2 \to \frac{(\bar{X}-\mu)^2}{\sigma^2/N}$ and $t \to \frac{\bar{X}-\mu}{\sigma/\sqrt{N}}$.

Hence, the ultimate mean and variance of t are

$$\text{Ult } E(t) = \frac{E\,(\bar{X}-\mu)}{\sigma/\sqrt{N}} = \frac{0}{\sigma/\sqrt{N}} = 0$$

$$\text{Ult } Var(t) = \frac{Var\,(\bar{X}-\mu)}{\sigma^2/N} = \frac{\sigma^2/N}{\sigma^2/N} = 1$$

so that irrespective of whether or not the normality assumption is met, when sample size becomes infinite the one-sample t statistic has its normal-theory mean and variance.