1-1-1973

On the robustness of certain tests for homogeneity of variance.

John v. S. Lochhead
University of Massachusetts Amherst

ON THE ROBUSTNESS OF CERTAIN
TESTS FOR HOMOGENEITY OF VARIANCE

A Dissertation Presented

By

JOHN v. S. LOCHHEAD

Submitted to the Graduate School
of the University of Massachusetts
in partial fulfillment of the requirements
for the degree of

DOCTOR OF EDUCATION

February 1973

Major Subject: Educational Statistics

ON THE ROBUSTNESS OF CERTAIN
TESTS FOR HOMOGENEITY OF VARIANCE

A Dissertation

By

JOHN v. S. LOCHHEAD

Approved as to style and content by:

(Chairman of Committee)

(Dean's Representative)

(Member)

(Member)

(Dean)

December 1972

On the Robustness of Certain Tests for Homogeneity of Variance. (February 1973)

John v. S. Lochhead, B. A., Johns Hopkins University
M.S., University of Massachusetts
Directed by: Dr. William C. Wolf

Abstract

The robustness of the following four tests for homogeneity of variance was investigated: Cochran's, Hartley's, Miller's Jackknife and Scheffe's. Both Type I and Type II error was determined under conditions of non-normal data and unequal sample sizes. All tests were on two samples and sample sizes ranged from 10 to 30. The results are based on Monte Carlo calculations of one thousand points. Cochran's, Hartley's and Miller's tests were found to perform well. Scheffe's test had poor power, but this may have been caused by the number of samples and their small size.

ACKNOWLEDGEMENTS

The author would like to express his thanks and appreciation to the following people for their contribution to the completion of this thesis.

To Betty Proper, for help with the literature search.

To the CDC 3600, without whose untiring efforts this project would have been impossible; and to the University Computing Center for a generous grant of computing time.

To Dr. Jimmie C. Fortune and Dr. William C. Wolf, for their advice, support, and careful reading of the manuscript.

To Dr. Robert Kleyle, for his advice, support, careful reading of the manuscript and for his assistance with the more mathematical sections of the dissertation, particularly Appendix II.

To Dr. Richard Kofler and Dr. Klaus Schultz, for their assistance with the less mathematical sections of Appendix II.

To Danielle G. Lochhead for her advice, support, careful reading of the manuscripts, and for her assistance in every section of the thesis, from the mathematics to the computer programming.

To Patricia Joiner, for typing the manuscript.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

Statistical Tests

Statistical tests form the basis of much of the "hard" research in education. The principle on which these tests operate is the following. Given some data, for example student test scores, one views it as a sample from all the possible data that might have been collected, i.e. the scores from all students that could have been tested. (Properly the population must be determined first and then a random sample taken from it. But in practice this is often not feasible and as a result the randomness of the sample is open to doubt.) The goal is to extract from the sample evidence on which to make inferences about the larger parent distribution. If the data is from two or more distinguishable groups and the aim is to infer that the groups are in some respect different, then a statistical test is performed. Each group of data is viewed as a sample from a parent population and the samples are compared, not to find out if there are differences in the samples, but rather to find if it is probable that there are differences between the populations from which the samples were supposedly drawn.

Parent populations are distributions whose mean and other moments may be unknown. Statistical tests are usually designed to distinguish these distributions on the basis of probable differences in their means or occasionally on differences in a specific higher moment. In the standard controlled experiment a hypothesis has been formulated which will indicate what changes the treatment will effect on the parent population. To validate the hypothesis it must be shown that the distributions differ in a particular characteristic, for example the mean, and not just in some unknown combination of mean and moments. While statistical tests usually claim to be specific to a particular factor, many respond to additional moments. In particular some tests for differences in means also measure differences in variance and may upon occasion report two populations different when their means are identical but their variances are not. Two of the most popular tests in educational research, the analysis of variance (ANOVA) and the T test, are in this category. For such tests it is necessary to be certain that all the populations used in the study have the same variance. This is usually done by assumption; but it may also be checked through a preliminary test that compares sample variances.

Test Assumptions and Robustness

The assumptions behind a test are carefully delineated by the statistician and then frequently ignored by the experimentor. There are two situations in which the assumptions may be violated without destroying the validity of the test The first is when the experimenter has reason to believe that the violation is of minor proportion, such as in the case of the assumption of random samples where carefully chosen nonrandom samples may differ very little from truly random ones. The other is when the test is robust to violation of the assumption.

> The term "robustness" is used to refer to the insensitivity of a statistical test to one or more of the assumptions underlying the test. The term is often used qualitatively, and the conditions under which a test exhibits a specified amount of robustness is stated vaguely or not at all. Bradley (1963, 1964) has taken issue with these practices, demonstrating very clearly that the degree to which a statistical test is robust depends on a rather large number of factors. In the light of Bradley's work, it is clear that any statement about robustness should include the conditions under which the test is said to be robust and some indication of the actual Type I and II errors.
>
> (Donaldson 1968)

The realities of experimental work often force the researcher to choose between using a test in violation of its assumptions or using no test at all. To make this decision intelligently it is essential that some knowledge of test robustness be available. The purpose of this study is to investigate the robustness of some of the tests for equality of variance. At present, little is known about this area and most competent researchers recommend that the tests be used only when the experimenter is confident that all the test assumptions are met. (See, for example, Box 1953.)

Tests for Homogeneity of Variance

The statistical tests used most frequently in educational research, the T test and ANOVA, are tests on means; both assume the equality of population variances. This is known as the assumption of homogeneity of variance; a detailed account of the importance of this assumption to ANOVA will be given in the next section. After the mean the most interesting population parameter is the second moment about the mean, i.e. the variance. Tests on this parameter are known as tests for homogeneity of variance; currently, they are rarely used by themselves and are mainly of interest as preliminary tests for ANOVA.

> Three instances in which testing for heterogeneity of population variances is worthwhile come to mind: a) when one wishes to make inferences about population variances because they are of scientific interest, b) when one suspects heterogeneity of variance in an analysis of variance in which not all of the factors have fixed effects, c) when one suspects heterogeneity of variances in a fixed effects analysis of variance in which the numbers of observations in the groups are widely disparate.
> (Glass 1965)

Unfortunately the assumptions underlying most tests for homogeneity of variance are inconsistent with their use as preliminary tests. To understand why this is so it is necessary to review the situations in which ANOVA is not robust to violation of the homogeneity of variance assumption; that is, those situations in which a preliminary test can be of value.

The Robustness of ANOVA

Homogeneity of variance is one of the basic assumptions underlying the use of ANOVA as a test for differences in means. Yet in many cases ANOVA is unaffected by violation of this assumption. There is an extensive literature describing the effects on ANOVA of violating each of its assumptions but much of this robustness literature does not meet the standards given by Donaldson (1968). In particular little attention is given to the subject of the robustness of Type II error rates. This is clearly seen, for example, in a paper by Box (1954) which is one of the standard references on the robustness of ANOVA to violation of the homogeneity of variance assumption. The paper briefly refers to the existence of a few power studies (Type II error) but includes data only on shifts in the observed percentile levels (Type I error). Yet most discussions of the homogeneity of variance problem (see for example Winer p. 92) are based on the results of Box or similar papers such as Welch (1938) and Newton (Lindquist, 1953).

The Type II error robustness studies that have been conducted (see Tang (1938), Hsu (1938), Hsu (1941) and Gronow (1951)) indicate that the power of ANOVA is robust to violation of the homogeneity of variance assumption. Studies by Davids and Johnson (1951), Donaldson (1968) and Clynch and Myers (Myers, 1971) have considered both Type I and Type II error rates. These papers conclude that observed percentile changes are the limiting factor in the robustness of ANOVA and that Type II error effects can be safely ignored.

Thus discusions on the robustness of ANOVA that are based soley on Type I error information do in this case reach valid conclusions. Table I summarizes the results of the Type I error studies. The table entries are the observed Type I error probabilities at the 95^{th} percentile. If ANOVA were completely robust all the entries would be .05; this, in fact, is the case if R, the ratio of sample sizes, is one or if V, the ratio of variances, is one. The latter case is to be expected since equality of sample size is not one of the assumptions of ANOVA and V=1 correspondes to there being no violation of assumptions. But the extreme robustness of ANOVA to violation of the homogeneity of variance assumption ($V\neq1$) under the condition of equal sample size (R=1) is surprising. This robustness deteriorates rapidly if $R\neq1$ as the bottom left and right corners of the table show. Errors in the right half of the table would appear to create a conservative test; but, since it is rarely known which sample represents the population of larger variance there is no way of determining in which half of the table a particular experiment will lie. Errors in the left half create liberal tests. In summary ANOVA is robust to violations of homogeneity of variance if sample sizes are equal.

(see Scheffé p. 340)

Table I

Therefore a preliminary test of variance is of value only in cases where the sample sizes are unequal.

One more robustness characteristic of ANOVA is of interest to this study. General discussions on the robustness of ANOVA, such as Eisenhart (1947), Cochran (1947) and Scheffé (1959), show that the test is remarkably robust to violations of the normality assumption. Thus a preliminary test that is sensitive to this characteristic may reject data that is acceptable to ANOVA. Box (1953) compared the use of a normality sensitive preliminary test to "putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port!")

The Robustness of Tests for Homogeneity of Variance

The implication of ANOVA's robustness characteristics to preliminary tests is that: 1) they must also be robust to deviations from the normality assumption since if they are not they may reject data that is perfectly acceptable to ANOVA and 2) tests for homogeneity of variance need only be employed when samples are of varying size. Thus if a test for homogeneity of variance assumes normal data and equal sample sizes, as most do, then it is a useful preliminary test for ANOVA only when it is robust to violation of these two assumptions. At present there is little to indicate that any of these tests have such robustness characteristics and the safest procedure is to avoid their use as preliminary tests.

Faced with this situation two solutions are possible: 1) create a new test whose assumptions are consistent with the conditions under which it is to be used: or 2) investigate whether any of the existing tests are robust with respect to violation of the troublesome assumptions. The latter course was chosen for this study. Four tests of homogeneity of variance were investigated; two that held special promise as tests robust to violation of their normality assumption and two that held less promise but were of interest because of their popularity. These tests were subjected to a Monte Carlo analysis (see chapter III) of their robustness with respect to two factors nonnormality and unequal sample size. No formal hypotheses were formulated because the point of the study was not to validate a theory but rather to discover whether or not any of the four tests were usable as preliminary tests. There was, however, an expectation that Cochran's and Hartley's tests would be shown to be inadequate and that the evidence from this report would discourage their use in the future. This prediction was found to be false.

CHAPTER II

REVIEW OF THE LITERATURE

The following tests of homogeneity of variance have appeared in the literature: Cochran (1941), Hartley (1950), Bartlett (1937), Wald (1947), Box-Andersen (1955), Levene (1960), Miller (1968), Moses (1963), Scheffé (1959), and Bartlett-Kendall (1946).[1] Box (1953) showed that the M_1 statistic upon which the Bartlett test is based is distributed as $(1+(1/2)\gamma_2)\chi^2_{k-1}$ rather than as χ^2_{k-1}; γ_2 is a measure of Kurtosis and thus the test is sensitive to departures from normality. Box (1955) further argued that the tests of Bartlett, Cochran, Hartley and Wald are sensitive to non-normality because they "tacitly compare some measure of variation among the variances with a theoretical value which is correct only for the normal distribution." Durand (1969) compared the robustness and power of Levene's test with Cochran's, Hartley's, and Bartlett's; he found that for small samples Levene's test did not have an advantage over the other two.

---

[1]The Bartlett-Kendall test is a special case of Scheffé's test.

Miller (1968) compared the power of the following tests: F, Box-Andersen, Jackknife with k=5[2], Jackknife with k=1, Levene, Box, and Moses. He also analysed their robustness with respect to violations from normality (see tables II & III) and drew the following conclusions.

> i) The F test is extremely non-robust
> ii) Box-Andersen and the Jackknife (k=1) are about equally powerful...
> iii) The observed significance levels under the null hypothesis for the Jackknife and Box-Andersen are more sensitive for small samples to the form of the distributions than in the case of larger sample sizes.
> iv) The Leven s test is quite robust, but lags far behind the Jackknife and Box-Andersen in power.
> (Miller, 1968)

Unfortunately all Miller's data is for samples of equal size.

The evidence to date indicates that only three tests are likely candidates in the search for a test robust to deviations from normality. These are the Scheffé, the Miller Jackknife and the Box-Andersen. All of these tests are difficult to compute. Furthermore, Scheffé does not state how samples should be divided into the subsamples needed for his test; nor has there been a study that compares the various possible methods.

---

[2]k defines the manner in which samples are divided into subsamples; specifically, it is the size of the subsamples.

(see Miller 1968)

Table II

(see Miller 1968)

Table III

This study expands on the present literature in three ways: 1) it considers the problem of unequal sample sizes that Miller ignored; 2) it investigates the issue of how to subdivide samples for Scheffé's test; and 3) it examines three tests that have so far received little attention in the robustness literature. These are the Scheffé, a test that is claimed to be robust to deviations from normality, plus Cochran's and Hartley's tests both of which are popular but not necessarily very robust. Techniques used for the study were not derived from any particular previous investigation but do share a great deal in common with some of those used by Miller (1968). As a result the data given in Chapter IV, tables XI to XV, can be easily compared to Miller's results given in tables II and III. Chapter III will describe the procedures used in this study and compare them to those of Miller.

CHAPTER III

PROCEDURE

The Range of Investigation

The following five tests were included in this study: Cochran's, Hartley's, Miller's, Scheffe with $J_i=2\nu_i$ and Scheffe with $J_i=5\nu_i$ (see appendix III). The Box-Andersen test was not included for two reasons: 1) Miller included it in his study; 2) it does not submit to the type of analysis used in this investigation. The Box-Andersen test is a corrected F test; but rather than altering the F distribution the correction alters the degrees of freedom. This means that corresponding to any particular test configuration there are a great many test statistic distributions. To experimentally generate each of these distributions would have been a very tedious and expensive task.

The investigation was simplified by considering only the case of two samples while sample sizes were restricted to the values 10, 20, 30. It was felt that considering tests on three or more samples would only complicate the issue without adding substantially to our understanding of the behavior of these tests. Small samples were used because they are the ones most frequently encountered in educational research and because tests on them are far more sensitive to violation of assumptions. Both simplifications also had the virtue of conserving computer time. (Despite the simplifications, this study was far from thrifty, about ten hours of computer time were used at a cost of $2,500.) On the other hand, the restricted range of the study does not give the Scheffé test a fair trial because it was designed primarily for use with more than two samples as well as for larger ones.

The original plan had been to investigate the effects of non-normality by independently varying the skewness and the kurtosis over a range of values; however, these two measures are not completely independent. It is not possible to obtain highly skewed distributions with normal values for kurtosis (see appendix I). Within this limitation one can obtain a wide range of distributions by using the property of linear additivity of moments. From a distribution f with moments μ_1, μ_2 ... and a distribution f' with moments μ'_1, μ'_2 ... one can construct a distribution

$$F = \frac{nf + n'f'}{n + n'}$$

whose moments are

$$\frac{n\mu_1 + n'\mu'_1}{n + n'}, \quad \frac{n\mu_2 + n'\mu'_2}{n + n'} \quad \ldots$$

Since the mean, variance, skewness and kurtosis are simple combinations of the first four moments it is possible to obtain a set of distributions with smoothly varying values of these parameters. Kurtosis, for example, can be varried over the values 2, 3, 4, 5, 6 while the skewness variance and mean are held constant. Skewness, on the other hand, cannot be varied much from zero unless the kurtosis is maintained at a high value. If Kurtosis is held at six then skewness can be varied between roughly 0 and 2. It turns out that there exists a set of common distributions whose parameters closely resemble some of those given above (see table IV). The kurtosis values are 1.8, 3, 6; those for skewness are 0 and 1.3. Two reasons for using this set are simplicity and the fact that Miller used it; thus any investigation doing likewise can be easily compared to his work. The chief disadvantage is that it is difficult to tell exactly what contributions the kurtosis and skewness make to the effects caused by non-normality. Since this investigation was interested in roughly determining the magnitude of the non-normality effects the above disadvantage was discounted.

Distribution Functions Used in the Study

	Domain	Probability Density Function	Mean	Variance	Skewness	Kurtosis
Rectangular	$\lvert x - m \rvert \leq h/2$	$1/h$	m	$h^2/12$	0	1.8
Normal	$-\infty < x < \infty$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$	m	σ^2	0	3
Laplace	$-\infty < x < \infty$	$\frac{1}{2\beta} e^{-\left\lvert\frac{x-\alpha}{\beta}\right\rvert}$	α	$2\beta^2$	0	6
Doubly Exponential	$-\infty < x < \infty$	$\frac{1}{\beta} \exp\left(-y - e^{-y}\right)$, $y = \frac{x-\alpha}{\beta}$	$\alpha + \gamma\beta$	$(\pi\beta)^2/6$	1.3	5.4

From Handbook of Mathematical Functions p. 930

Table IV

The Monte Carlo Calculation

The term Monte Carlo calculation is used to refer to any experiment where the data are synthetically generated through some sort of a random number generator rather than by real world events. With this technique it is possible to submit to a statistical test samples that come from populations whose parameters are known, in fact determined, by the experimenter. Repeating this process many times creates a histogram of the statistical test results which is an experimentally generated test statistic distribution. This distribution may be compared with the theoretical statistic distribution to determine how certain parameters in the population from which the samples were drawn effect the test statistic. In Miller's work the statistic distribution was never explicitly displayed; instead each test result was compared with a particular percentile in the theoretical distribution and the number exceeding that value was recorded. In this study the computer output included both a histogram of the test statistic distribution and a display of information calculated from the histogram. The nature of this information will be explained in the next section. The advantages of plotting the histogram will be discussed in chapter V.

The Computer Program

The data presented in this study were generated with a computer program written to investigate robustness problems by means of a Monte Carlo calculation. The main program (Program Robust) utilizes various subroutines and calculates basic sample statistics. By changing the subroutines it is possible to investigate the robustness of any test and under any set of conditions for which subroutines can be written. Thus to investigate the robustness of Cochran's test for data from a Laplacian distribution it is necessary to have two subroutines, one to calculate Cochran's C statistic and one to generate random data from a Laplacian distribution. The output of Program Robust is a histogram plot of the test statistic distribution. For the above example the histogram would be a frequency distribution of Cochran's C statistic, calculated from Laplacian data, with sample sizes determined in the data cards. From the histogram it is possible to calculate several items of interest. The $(1-\alpha)$ percentile is calculated by finding a point on the test statistic axis above which α% of the points lie. A confidence interval about this point can also be calculated by counting a certain number of points to each side of the experimentally determined critical value. (See appendix II). When Program Robust is used to calculate power the reverse of the above procedure is employed. A point on the test statistic axis is marked corresponding to the percentile found for the null hypothesis distribution and the number of points lying above that point is recorded. All of the above measures can be performed automatically by the computer program.

Program Robust is composed of five sections: 1) The first section reads in data cards which include information on the number and size of samples. 2) The second section determines the range of the histograms; it must be altered if the number of histograms is changed or if the range of a histogram is altered. 3) The next section calls data from the distribution program and calculates the basic statistics of each sample. The call statement is changed whenever a different distribution is to be investigated. 4) The fourth section calls the testing programs (Cochran, Hartley, etc.) and enters their results in the appropriate histograms. This section requires extensive revision each time a test is added or dropped from the program. 5) The final section calls the plot function, plots the histograms, and calls the programs that calculate statistics on these histograms (i.e. percentiles and Type II error rates). This section requires revision if the number of histograms is altered or if the histogram parameters are changed.

Distribution Subroutines

Most computer systems include a random number generator that gives sample points from a uniform distribution over the range 0. - 1.. There are several techniques for generating other distributions from this given one. Uniform distributions of different variance and mean can be generated by multiplication and subtraction.

$$x' = ax - b$$

Normally distributed random variables can be generated by using the following rule. The sum of n uniformly distributed numbers (on the interval 0. - 1.) behaves like a number from a normal population of mean n/2 and variance n/12. Thus to generate the population N(0,1)

$$x' = \sum_{i=1}^{12} x_i - 6$$

(see Tocher, 1963)

For the other distributions the acceptance rejection technique was employed. (See Handbook of Mathematical Functions p.952). This requires that the distribution's density function f(x) be in a calculable form and that its maximum F be known. Two uniform random numbers are generated, x_1 and x_2, x_1 is transformed to x_1', a variable whose range is large enough to cover at least 99% of the f(x) distribution and $f(x_1')$ is calculated. If $x_2 < f(x_1')/F$ then $f(x_1')$ is accepted as a random deviate from the f distribution; otherwise it is not and the process is repeated until an acceptable number is obtained.

In every case once a program had been written to produce a given distribution it was checked by generating an emperical distribution of 1,000 points. The mean, variance, skewness and kurtosis of this distribution was calculated and compared with the expected values. Absolute agreement was neither anticipated nor found; because 1) the distribution generation techniques are approximate and 2) even for samples of size 1,000, n is too small to expect perfect agreement. Nevertheless, only three of the values in table V differ significantly from the expected values given in table IV; these are the kurtosis for the Laplacian and the doubly exponential distributions and the skewness for the doubly exponential distribution. With the information presently available it is impossible to tell whether the values in table V are better estimators of the computer generated distribution parameters than those given in table IV. This uncertainty does limit the interpretations that can be drawn from this study but it does not effect the conclusions given in Chapter V.

The Test Programs

The programs written to calculate Cochran's, Hartley's, Miller's and Scheffé's tests were tested by solving problems with precalculated answers and by comparing distribution generated percentiles with those given in the standard tables. Three comparisons are possible from the data included in tables VI - X (the top three cells of row two); several other cases using more than two samples were also tested.

Experimentally Generated Values

for the Parameters of the Four

Distributions used in the Study

Type of Distribution	Mean	Variance	Skewness	Kurtosis
Laplace	.020	.952	.035	5.076
Doubly Exponential	.014	1.027	.903	3.794
Normal	-.008	0.999	-0.038	3.037
Rectangular	-0.012	0.987	-0.005	1.787

Table V

CHAPTER IV

DATA

The Monte Carlo calculations produced over 250 graphs a few of which have been included in this paper (see figures 1 - 8). Approximately half of these graphs were statistic distributions with valid null hypotheses; that is distributions generated by samples from populations of equal variance. The rest of the distributions were generated by samples from populations of unequal variance.

This chapter will start by using the equal variance data to investigate Type I error robustness. Tables VI to X report observed percentiles and show how these change as a function of non-normality and as a function of varying sample size. Next the robustness of Type II error is considered. Tables XI to XV give values for power (under an alternative hypothesis $\sigma_2^2=5$) and show how power is effected by non-normal data and by unequal sample sizes. At the end of the chapter a few crude power curves are presented and some of the statistic distributions are examined to show how they relate to the tables.

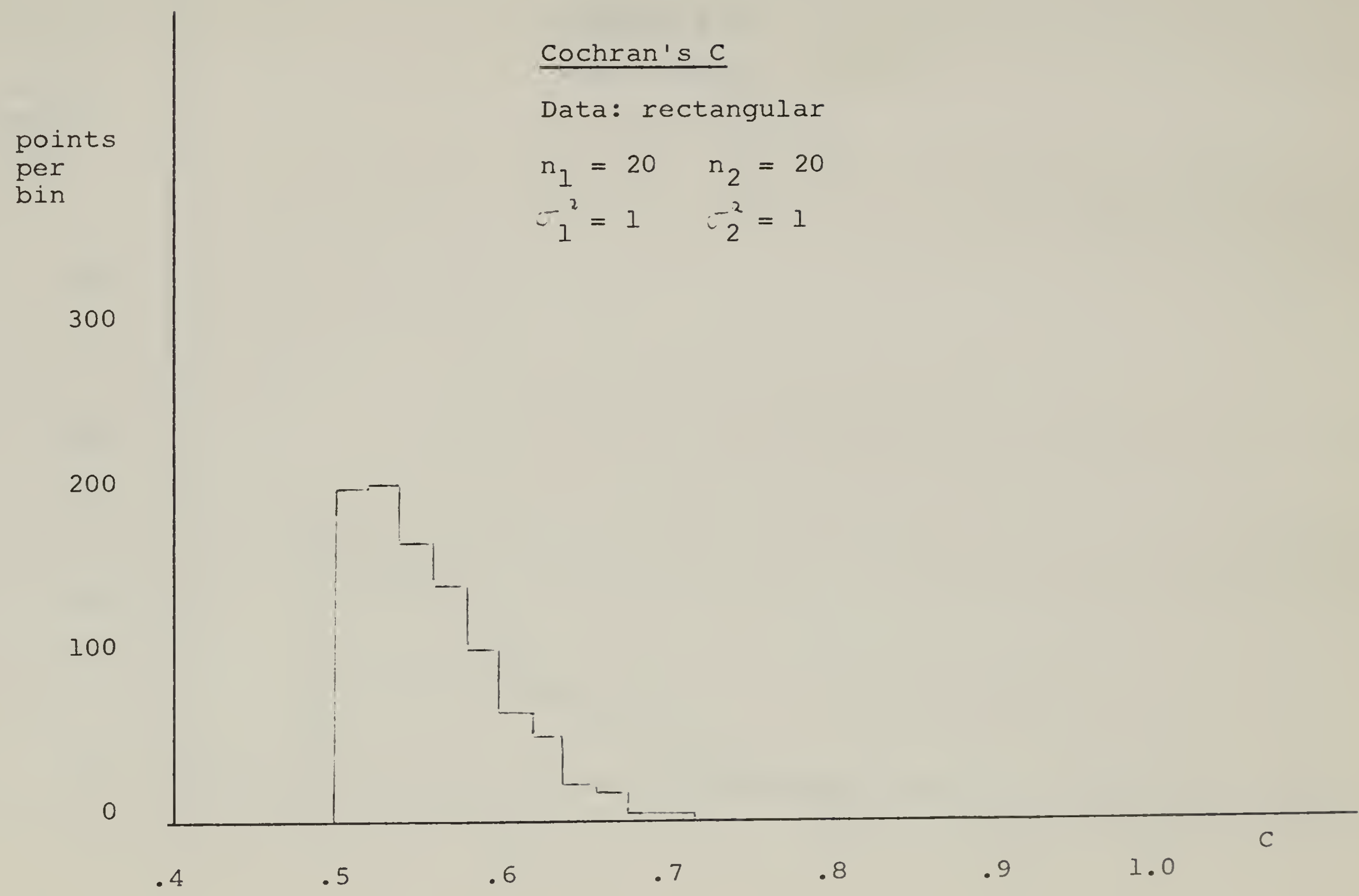

Cochran's C
Data: rectangular
n1 = 20 n2 = 20
σ1² = 1 σ2² = 1
points per bin
300
200
100
0
.4
.5
.6
.7
.8
.9
1.0
C

Figure 1

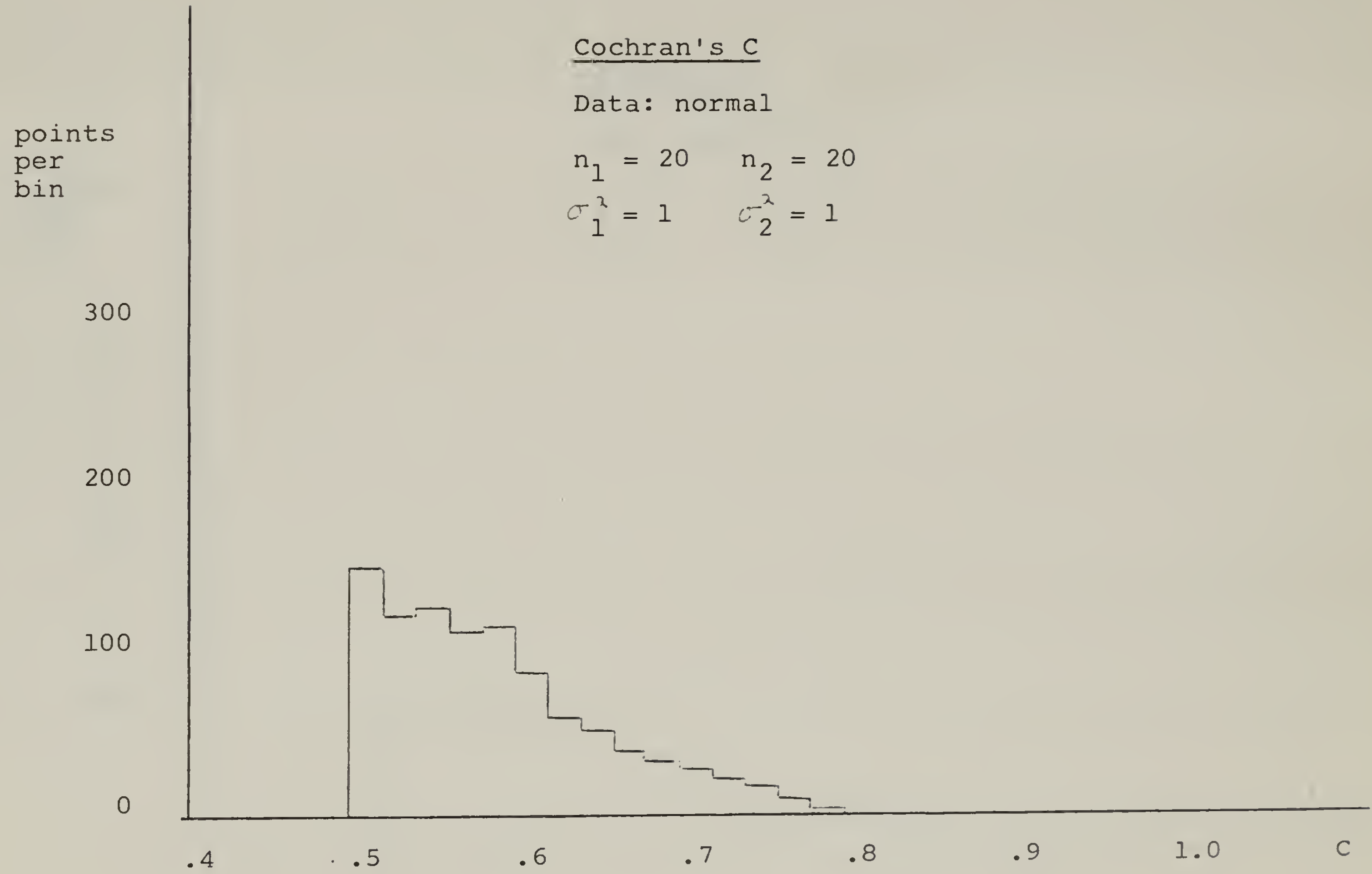

Cochran's C
Data: normal
$n_1 = 20$ $n_2 = 20$
$\sigma_1^2 = 1$ $\sigma_2^2 = 1$
points per bin
300
200
100
0
.4
.5
.6
.7
.8
.9
1.0
C

Figure 2

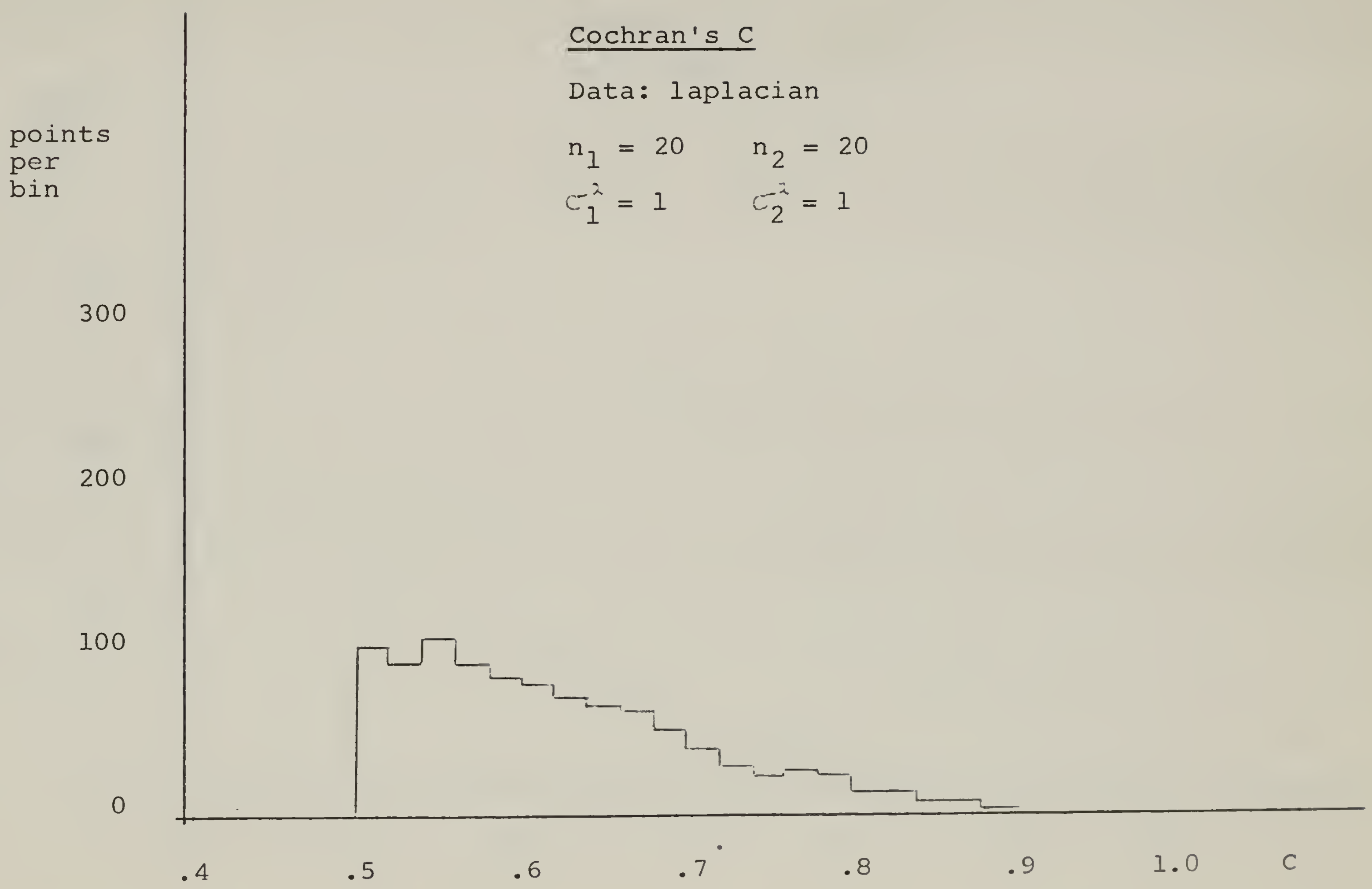

Cochran's C
Data: laplacian
$n_1 = 20$
$n_2 = 20$
$\sigma_1^2 = 1$
$\sigma_2^2 = 1$
points per bin
300
200
100
0
.4
.5
.6
.7
.8
.9
1.0
C

Figure 3

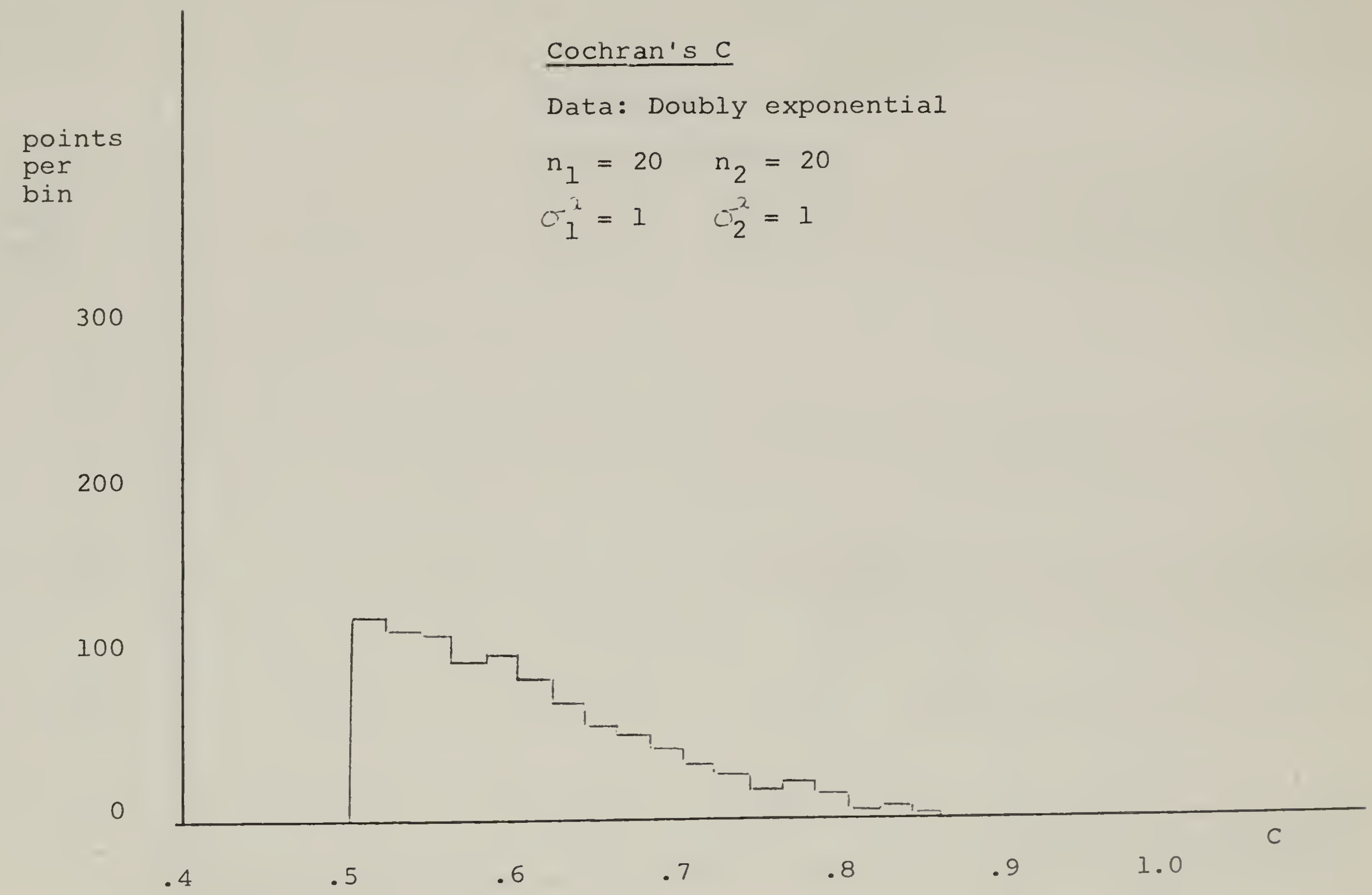

Cochran's C
Data: Doubly exponential
$n_1 = 20$ $n_2 = 20$
$\sigma_1^2 = 1$ $\sigma_2^2 = 1$
points per bin
300
200
100
0
.4
.5
.6
.7
.8
.9
1.0
C

Figure 4

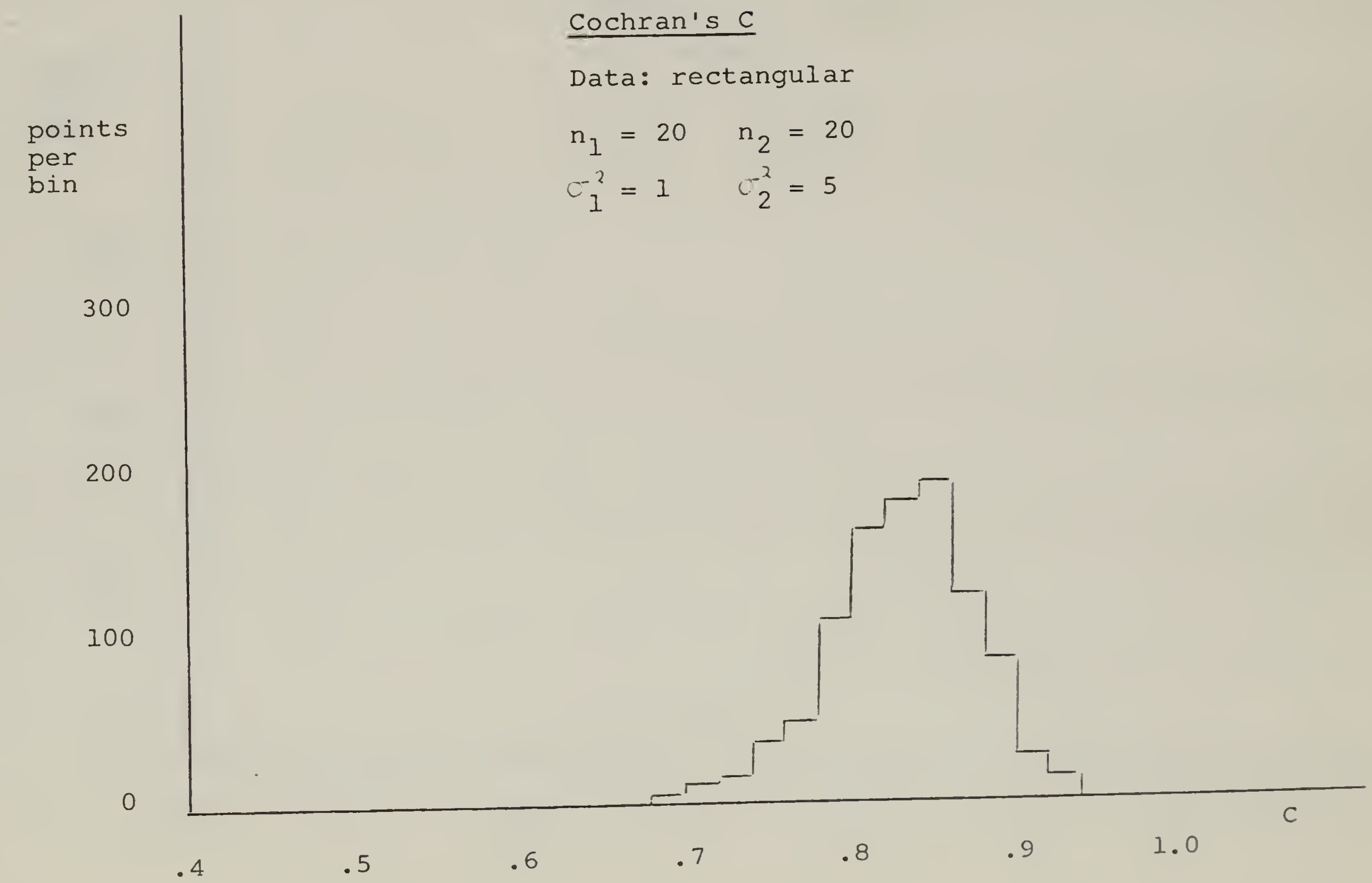

Cochran's C
Data: rectangular
$n_1 = 20$ $n_2 = 20$
$\sigma_1^2 = 1$ $\sigma_2^2 = 5$
points per bin
300
200
100
0
.4
.5
.6
.7
.8
.9
1.0
C

Figure 5

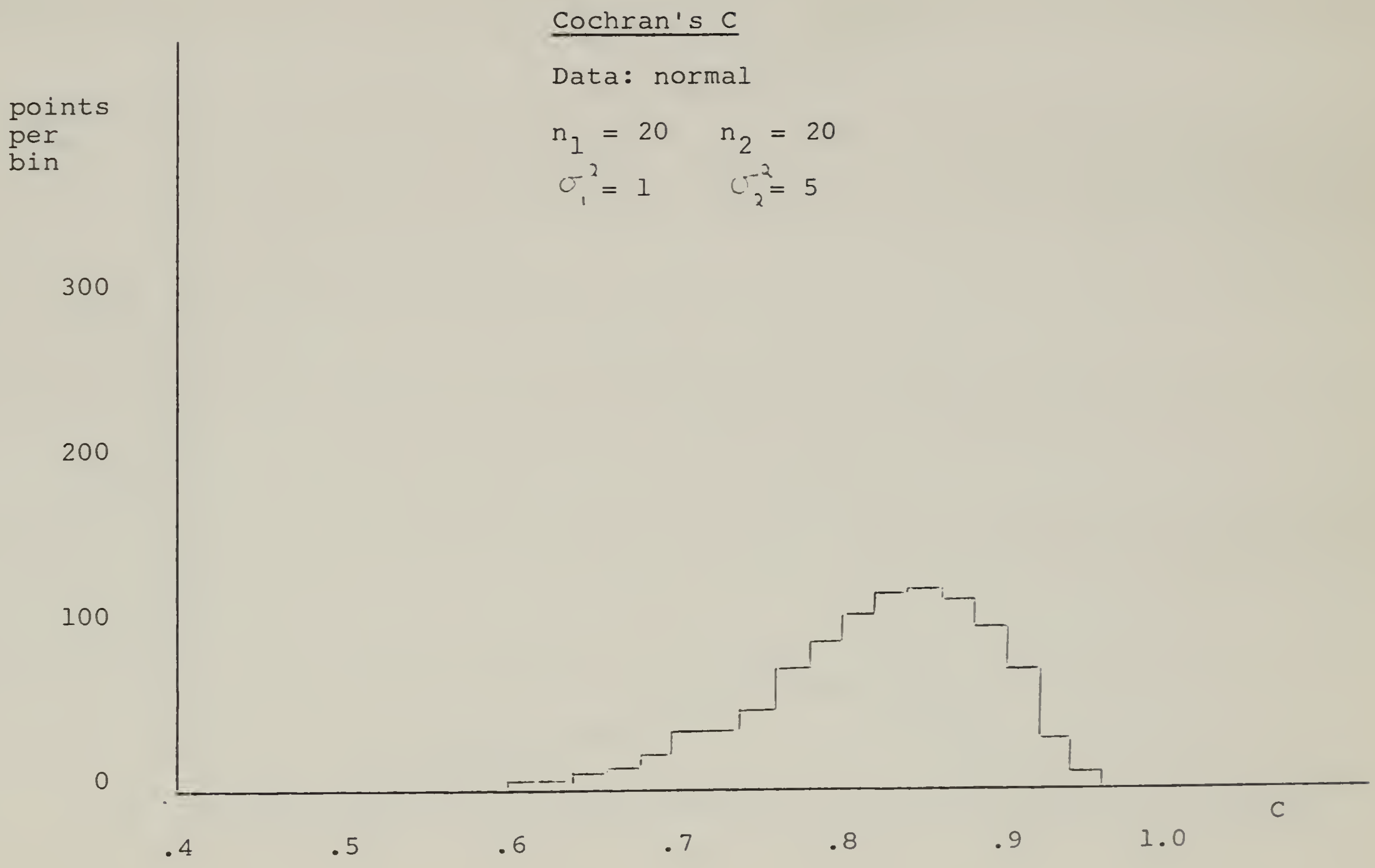

Cochran's C
Data: normal
n1 = 20 n2 = 20
σ1² = 1 σ2² = 5
points per bin
300
200
100
0
.4
.5
.6
.7
.8
.9
1.0
C

Figure 6

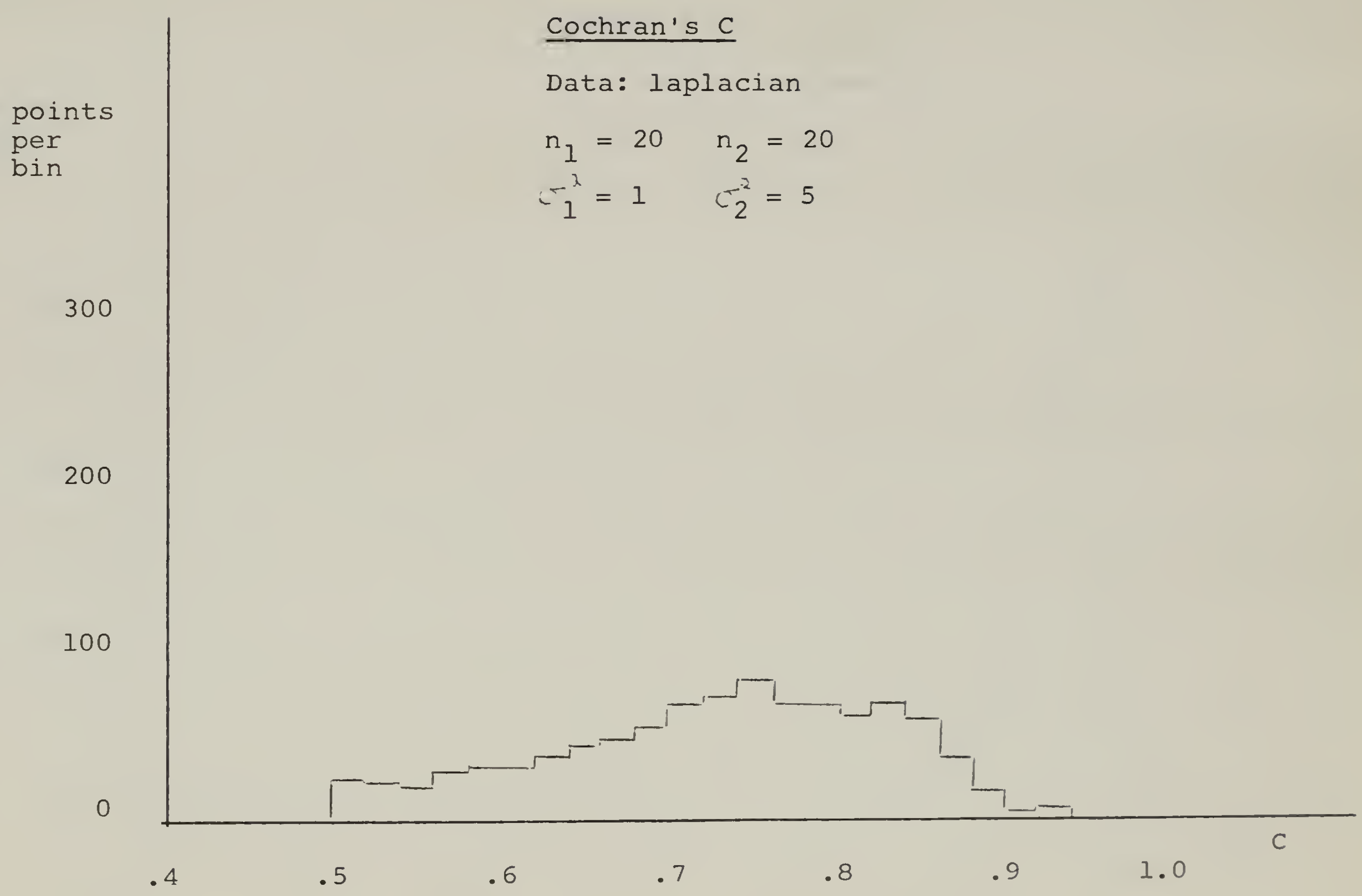

Cochran's C
Data: laplacian
$n_1 = 20$ $n_2 = 20$
$\sigma_1^2 = 1$ $\sigma_2^2 = 5$
points per bin
300
200
100
0
.4
.5
.6
.7
.8
.9
1.0
C

Figure 7

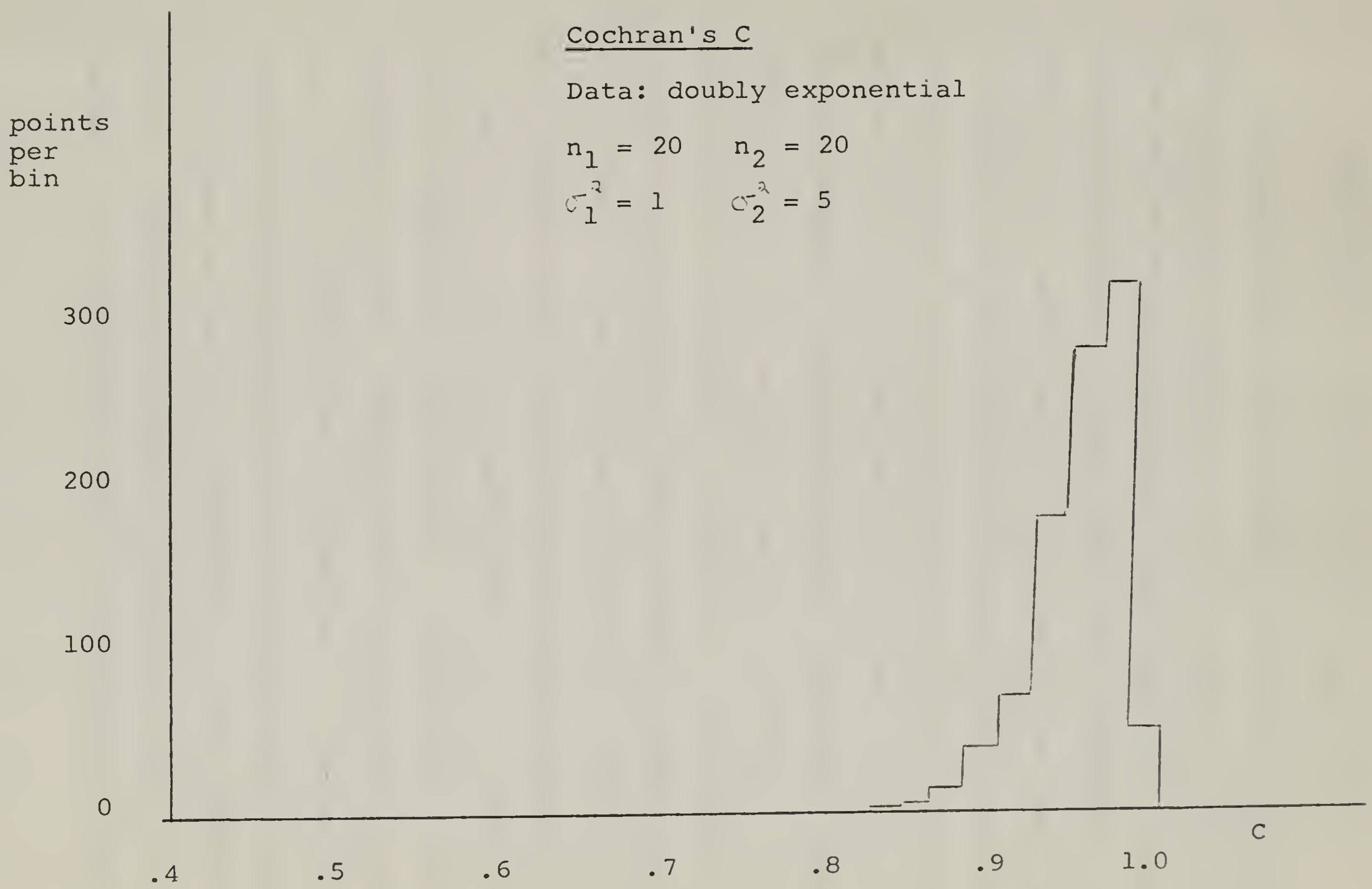

Cochran's C
Data: doubly exponential
n1 = 20   n2 = 20
σ1² = 1   σ2² = 5
points per bin
300
200
100
0
.4
.5
.6
.7
.8
.9
1.0
C

Figure 8

Robustness of the Type I Error Probability

The valid null distributions were used to determine percentiles (see tables VI - X). By comparing the percentiles generated under different conditions it is possible to gain some measure of the Type I error robustness. When examining the tables notice first that column two presents the percentiles generated from normally distributed data. The first three cells of this column are for data consistent with all the assumptions of the test and can be directly compared to the values given in standard tables. Reading across the table one can see the effect of violating the normality assumption. Columns 1, 2 and 3 permit a comparison of the effect of three different values for kurtosis while columns 2 and 4 allow a check on the effect of skewness. Reading down the table reveals the effects of changes in sample size; with the second half of the table containing data on samples of unequal size. For Cochran's, Hartley's and Miller's tests, percentiles are a function of sample size. For Scheffé's test the degrees of freedom are determined by the manner in which the samples are divided into subsamples and are independent of the original sample sizes. Therefore no change in the percentiles is to be expected while progressing down the two Scheffé tables.

Observed Percentiles for Cochran's Test

$\sigma_1^2 = 1 \quad \sigma_2^2 = 1$

N_1 N_2	$1-\alpha$	$\gamma_1 = 0$ $\beta_2 = 1.8$		$\beta_2 = 3$		$\beta_2 = 6$		$\gamma_1 = 1.3$ $\beta_2 = 5.4$	
10 10	.99.	.794	.815 .784	.880	.920 .856	.907	.930 .895	.882	.896 .873
	.95	.728	.718 .740	.803	.814 .792	.854	.865 .845	.835	.846 .827
	.90	.692	.684 .702	.763	.774 .754	.818	.830 .810	.799	.812 .790
20 20	.99	.700	.720 .686	.768	.778 .760	.850	.870 .837	.816	.834 .799
	.95	.646	.658 .638	.725	.737 .716	.792	.806 .782	.763	.774 .751
	.90	.623	.630 .618	.691	.701 .682	.752	.765 .739	.719	.732 .710
30 30	.99	.660	.680 .652	.717	.733 .711	.797	.827 .783	.771	.792 .758
	.95	.621	.630 .616	.671	.681 .663	.743	.757 .731	.713	.728 .700
	.90	.604	.610 .599	.645	.653 .639	.706	.715 .697	.681	.690 .671
20 30	.99	.689	.710 .676	.754	.780 .744	.818	.845 .807	.789	.820 .778
	.95	.636	.695 .630	.703	.717 .693	.763	.774 .751	.730	.745 .718
	.90	.615	.620 .611	.669	.678 .661	.726	.735 .718	.696	.705 .689
10 20	.99	.755	.800 .737	.830	.853 .818	.908	.920 .897	.880	.900 .869
	.95	.693	.708 .608	.769	.780 .758	.840	.860 .830	.815	.830 .805
	.95	.655	.663 .649	.728	.739 .718	.801	.813 .788	.772	.783 .762
10 30	.99	.745	.770 .731	.820	.837 .795	.880	.900 .866	.849	.858 .842
	.95	.677	.689 .668	.746	.758 .735	.814	.829 .799	.788	.808 .776
	.90	.649	.656 .643	.711	.719 .704	.776	.786 .768	.754	.764 .743

Table VI

Observed Percentiles for Hartley's Test

$\sigma_1^2 = 1 \quad \sigma_2^2 = 1$

	$1-\alpha$	$\gamma_1 = 0$, $\beta_2 = 1.8$		$\gamma_1 = 0$, $\beta_2 = 3$		$\gamma_1 = 0$, $\beta_2 = 6$		$\gamma_1 = 1.3$, $\beta_2 = 5.4$	
10 10	.99	3.93	3.70 4.20	6.70	8.15 5.80	11.40	15.00 10.50	7.45	8.00 6.45
	.95	2.67	2.53 2.89	4.08	4.29 3.86	6.96	7.53 6.68	5.05	5.38 4.74
	.90	2.22	2.17 2.29	3.22	3.42 3.09	4.86	5.25 4.55	3.99	4.30 3.75
20 20	.99	2.33	2.57 2.22	3.30	3.53 3.16	5.60	6.45 5.05	4.18	5.10 3.90
	.95	1.83	1.93 1.77	2.63	2.80 2.53	3.68	4.10 3.59	3.22	3.40 3.02
	.90	1.65	1.69 1.62	2.22	2.33 2.14	3.04	3.28 2.83	2.57	2.74 2.44
30 30	.99	1.93	2.10 1.86	2.50	2.75 2.42	3.90	4.70 3.60	3.40	3.80 3.15
	.95	1.64	1.67 1.60	2.05	2.14 1.98	2.88	3.10 2.71	2.40	2.71 2.33
	.90	1.53	1.56 1.50	1.82	1.88 1.77	2.42	2.51 2.31	2.14	2.23 2.04
20 30	.99	2.18	2.40 2.07	2.94	3.50 2.86	4.50	5.50 4.20	3.70	4.40 3.45
	.95	1.76	1.84 1.70	2.36	2.52 2.25	3.20	3.45 3.00	2.68	2.90 2.56
	.90	1.59	1.64 1.57	2.02	2.10 1.95	2.64	2.78 2.53	2.30	2.42 2.22
10 20	.99	3.15	4.00 2.80	4.95	5.47 4.40	8.80	9.50 8.20	2.73	3.50 2.62
	.95	2.26	2.39 2.13	3.28	3.55 3.14	4.68	5.04 4.47	1.76	1.96 1.59
	.95	1.90	1.97 1.84	2.70	2.83 3.55	3.67	3.81 3.52	1.37	1.46 1.29
10 30	.99	2.90	3.85 2.68	4.40	5.10 3.87	7.20	8.00 6.30	5.60	6.15 5.30
	.95	2.10	2.24 2.01	2.90	3.12 2.78	4.37	4.80 4.07	3.68	4.10 3.45
	.90	1.85	1.91 1.80	2.45	2.54 2.38	3.49	3.78 3.36	3.05	3.24 2.90

Table VII

Observed Percentiles for Miller's Test

(based on the absolute value of Miller's T) $\sigma_1^2 = 1$ $\sigma_2^2 = 1$

N_1	N_2	$1-\alpha$	$\gamma_1 = 0$, $\beta_2 = 1.8$		$\gamma_1 = 0$, $\beta_2 = 3$		$\gamma_1 = 0$, $\beta_2 = 6$		$\gamma_1 = 1.3$, $\beta_2 = 5.4$	
10	10	.99	2.50	2.95 / 2.36	2.97	3.40 / 2.87	3.30	3.60 / 3.20	3.38	3.60 / 3.30
		.95	1.75	1.85 / 1.70	2.03	2.15 / 1.94	2.44	2.56 / 2.36	2.50	2.60 / 2.42
		.90	1.40	1.46 / 1.36	1.65	1.72 / 1.59	1.90	2.00 / 1.83	1.93	1.98 / 1.88
20	20	.99	2.45	2.57 / 2.35	2.85	3.05 / 2.77	2.85	3.00 / 2.70	3.20	3.55 / 2.97
		.95	1.77	1.84 / 1.73	2.08	2.18 / 2.04	2.12	2.20 / 2.06	2.31	2.39 / 2.23
		.90	1.45	1.50 / 1.41	1.76	1.83 / 1.70	1.75	1.81 / 1.70	1.91	1.95 / 1.82
30	30	.99	2.37	2.60 / 2.28	2.72	2.80 / 2.55	3.20	3.50 / 3.05	2.83	3.03 / 2.70
		.95	1.85	1.93 / 1.78	1.98	2.09 / 1.91	2.17	2.32 / 2.10	2.15	2.23 / 2.10
		.90	1.49	1.54 / 1.45	1.58	1.63 / 1.54	1.84	1.91 / 1.78	1.81	1.86 / 1.77
20	30	.99	2.33	2.50 / 2.23	2.63	3.00 / 2.55	2.87	3.05 / 2.75	3.00	3.30 / 2.80
		.95	1.80	1.85 / 1.76	2.03	2.12 / 1.94	2.21	2.32 / 2.14	2.16	2.24 / 2.08
		.90	1.52	1.57 / 1.47	1.70	1.74 / 1.66	1.83	1.89 / 1.77	1.78	1.85 / 1.73
10	20	.99	2.80	2.98 / 2.40	3.10	3.33 / 3.00	3.60	4.23 / 3.35	3.35	3.70 / 3.17
		.95	1.77	1.82 / 1.72	2.15	2.24 / 2.06	2.36	2.44 / 2.30	2.44	2.53 / 2.37
		.95	1.49	1.53 / 1.45	1.68	1.75 / 1.62	1.99	2.08 / 1.92	1.93	2.02 / 1.87
10	30	.99	2.80	3.00 / 2.70	2.80	3.00 / 2.70	3.50	4.03 / 3.37	3.97	4.20 / 3.80
		.95	1.88	2.04 / 1.82	2.08	2.15 / 2.00	2.39	2.55 / 2.33	2.30	2.53 / 2.24
		.90	1.49	1.56 / 1.44	1.66	1.71 / 1.61	1.91	1.98 / 1.85	1.90	1.98 / 1.83

Table VIII

Observed Percentiles for Scheffe's Test

with two Subsamples $\sigma_1^2 = 1$ $\sigma_2^2 = 1$

N_1 N_2	$1-\alpha$	$\gamma_1 = 0$									$\gamma_1 = 1.3$		
		$\beta_2 = 1.8$			$\beta_2 = 3$			$\beta_2 = 6$			$\beta_2 = 5.4$		
10 10	.99	83.00	100.00	72.00	100.		79.00	80.00	100.	69.00	90.00	100.	68.00
	.95	20.00	22.67	17.00	20.00	27.00	16.00	20.43	22.50	18.67	23.33	28.00	21.00
	.90	7.78	9.50	6.82	8.80	9.90	7.33	10.14	11.80	8.75	9.11	10.33	7.70
20 20	.99	100.			83.00	100.	66.00	100.		89.50	100.		
	.95	17.50	24.20	14.75	19.00	21.83	16.40	22.00	26.40	17.00	19.00	22.50	15.50
	.90	8.18	9.11	7.20	8.46	9.60	7.14	7.87	9.00	7.20	8.09	9.00	7.18
30 30	.99	81.00	100.	59.00	100.		83.00	65.00	100.	57.00	84.00	100.	72.00
	.95	18.50	23.00	15.75	18.00	24.00	15.00	15.00	18.00	13.88	18.33	22.75	15.33
	.90	8.33	9.50	7.00	7.82	9.29	7.24	8.09	9.00	7.18	7.70	8.47	6.82
20 30	.99	77.00	100.	61.00	82.00	100.	67.00	100.			80.00	100.	65.00
	.95	14.60	17.00	12.80	14.00	18.00	13.30	23.67	31.50	20.67	19.40	23.50	16.00
	.90	7.50	8.39	6.74	7.25	8.33	6.38	9.56	10.67	8.38	8.29	9.63	7.11
10 20	.99	41.00	51.50	38.00	100.		91.33	100.		71.00	69.00	94.00	60.00
	.95	13.50	15.00	12.40	21.29	24.00	17.25	19.2	22.33	16.67	17.75	24.67	15.00
	.90	7.59	8.18	7.00	8.89	10.00	7.88	7.29	8.00	6.54	7.07	7.79	6.55
10 30	.99	100.		76.00	87.00	100.	79.00	100.			100.		66.00
	.95	19.67	23.00	17.00	17.00	21.33	15.71	24.00	36.00	21.33	21.00	23.67	18.00
	.90	8.80	9.78	8.13	8.30	9.43	7.00	12.25	14.00	11.11	8.67	10.40	7.86

Table IX

Observed Percentiles for Scheffe's Test

with Five Subsamples $\sigma_1^2 = 1$ $\sigma_2^2 = 1$

		$\gamma_1 = 0$						$\gamma_1 = 1.3$	
N_1 N_2	$1-\alpha$	$\beta_2 = 1.8$		$\beta_2 = 3$		$\beta_2 = 6$		$\beta_2 = 5.4$	
10 10	.99	8.60	11.20 7.70	9.25	15.25 8.00	10.30	12.00 9.40	10.80	15.00 8.80
	.95	5.06	5.68 4.40	4.71	5.00 4.44	5.40	6.73 4.75	5.10	5.60 4.46
	.90	3.27	3.66 2.95	3.10	3.30 2.92	3.43	3.80 3.21	3.18	3.64 2.83
20 20	.99	13.17	13.50 12.50	12.60	15.00 10.00	10.80	15.00 9.80	11.40	14.40 9.30
	.95	5.17	5.50 4.85	6.10	7.08 5.30	5.40	6.60 4.97	5.50	6.35 4.80
	.90	3.58	3.77 3.36	3.68	4.08 3.34	3.72	4.17 3.35	3.48	4.00 3.00
30 30	.99	10.40	15.00 8.80	9.60	11.60 8.07	11.30	15.00 10.13	11.60	14.60 9.13
	.95	4.92	5.70 4.38	5.17	6.08 4.34	4.80	5.67 4.28	4.80	5.70 4.10
	.90	3.07	3.47 2.76	3.47	3.89 3.20	3.23	3.52 3.05	3.10	3.49 2.86
20 30	.99	11.50	15.00 10.60	10.60	15.00 9.40	11.40	15.00 9.40	11.00	15.00 9.30
	.95	5.70	6.75 4.94	5.30	6.70 4.56	5.77	6.40 5.37	5.80	7.20 5.45
	.90	3.74	4.20 3.38	3.31	3.83 3.00	3.65	4.28 3.20	3.59	4.12 3.35
10 20	.99	14.50	22.00 13.00	13.40	15.00 11.60	14.00	18.00 13.25	14.00	18.50 13.00
	.95	8.22	8.71 7.70	8.00	8.50 7.07	7.50	8.21 6.75	7.18	7.46 6.83
	.95	5.17	5.44 4.91	5.47	6.23 4.98	4.91	5.20 4.62	4.83	5.09 4.61
10 30	.99	20.00	26.25 19.17	17.17	18.00 16.50	15.00	20.00 14.75	15.00	19.50 14.00
	.95	10.10	10.50 9.75	9.60	10.56 8.63	10.60	11.25 10.13	8.78	9.20 8.42
	.90	7.76	8.14 7.31	5.91	6.25 5.59	7.00	7.50 6.55	6.25	6.58 5.92

Table X

Cochran's test. Examining the first row of table VI, one can see that the 95th percentile of the normal hypothesis has a value of 0.8 $\pm$ 0.01 which is close to the value given for the 99th percentile when β_2 = 1.8 (0.79) and just less than the value given for the 90th under β_2 = 6 (0.82). Thus if the value 0.8 is used in a situation where the data is rectangularly distributed the test will prove to be conservative; however, if the data is from a distribution with high kurtosis the test will be quite liberal. For cases in which the kurtosis is unknown a liberal test may be formed by using 0.8 as the 99th percentile and a conservative test made by using 0.8 as the 90th, or perhaps even 85th percentile.

If the experimenter has access to only the standard tables then he may estimate the effects of non-normal data by creating an uncertainty band about the table value of $\pm$ 5 percentiles. Or he may create a consistently conservative test by using the table value five percentiles below the percentile he desires. The above discussion just concerns the effect of kurtosis but since skewness (see column four) does not appear to effect the percentiles significantly the argument holds for non-normality in general. The 99th and 90th percentiles are similar to the 95th in robustness and may be handled in the same manner. Of course the upper end for the confidence interval of the 99th percentile cannot be determined by adding five percentiles, but for a conservative test only the lower end is important. Sample size and inequality of sample size do not effect the degree of robustness; the "rule" of ± 5 percentiles holds down the entire table.

Since the standard tables do not include the percentiles for samples of unequal size some rule has to be given for converting the unequal sample size case to an equal sample size situation.

From the data in table VI it would appear that the average sample size will result in a liberal test. The 95th percentile for the case $n_1 = 10$ $n_2 = 30$ is .75 which slightly exceeds the value 0.73 found for the 95th when $n_1 = 20$ $n_2 = 20$. On the other hand choosing the smallest sample size results in a conservative test, the 95th percentile for the case $n_1 = 10$ $n_2 = 10$ is 0.8. Since the above effects are less than those caused by violations of the normality assumption, they may be ignored in most cases, and either system may be used.

Hartley's test. The analysis of Hartley's test, table VII, is essentially the same as for Cochran's test. The effect of high kurtosis ($\beta_2 = 6$) is a bit larger than for Cochran's test, but skewness seems to decrease this effect. To form a conservative test it is probably necessary to use the 85th percentile instead of the 95th. There would appear to be some interaction between the use of differing sample sizes and violations from normality. The bottom half of the table, especially the case $n_1 = 10$ $n_2 = 20$, shows far more variation than does the top part (3.2, 5.0, 8.8, 2.7 for the 99th percentile).

The situation concerning a method for finding an equal sample size percentile appropriate for the unequal sample size case is the same as it was for Cochran's test.

Miller's test. Miller's test, table VII, would appear to be more robust than either Cochran's or Hartley's. Instead of shifting α levels by 5 or 10 percentiles a shift of 2 to 5 would seem adequate. However, unlike Cochran's and Hartley's tests, Miller's test is sensitive to skewness. As was the case with Cochran's test, unequal sample sizes do not affect the robustness to non-normality. Using the average sample size is definitely the proper preceedure for this test. The percentiles for the case $n_1 = 10$ $n_2 = 30$ and $n_1 = 20$ $n_2 = 20$ are identical within experimental error.

Scheffé's test. The two subsample version of Scheffé's test, table IX, is robust both to deviations from normality and to unequal sample size. Percentiles need not be changed by more than 1. The five subsample version, table X, is almost as robust but is less stable for situations with unequal sample sizes. The changes that do occur are true across the table, that is they are robust with respect to non-normality and are a function only of sample size.

If the experimenter has access to a table such as table X he should use the data given there rather than that in the standard tables. Thus the 95th percentile for the case $n_1 = 10$ $n_2 = 30$ would be 10. rather than 5..

Robustness of the Type II Error Probability

Although the invalid null distributions can be analysed in terms of percentiles it is usually more informative to use them for a measure of power. Tables XI to XV give power values for the particular alternative hypothesis $\sigma_1^2 = 1$ $\sigma_2^2 = 5$ (the null hypothesis being $\sigma_1^2 = 1$ $\sigma_2^2 = 1$). Since this alternative is quite far from the null, power ratings should be high. The numbers given in the tables XI to XV are the proportion of the alternative distribution that lies above the critical value for normally distributed data. An example is given in figure 9 where both the normal null hypothesis valid distribution and the laplacian null hypothesis invalid distribution are plotted together. The double cross hatched area is the 5% of the null valid distribution that falls above the C value 0.725. The single cross hatched area (including the double cross hatched) is the 56.5% of the laplacian distribution that falls above C = 0.725.

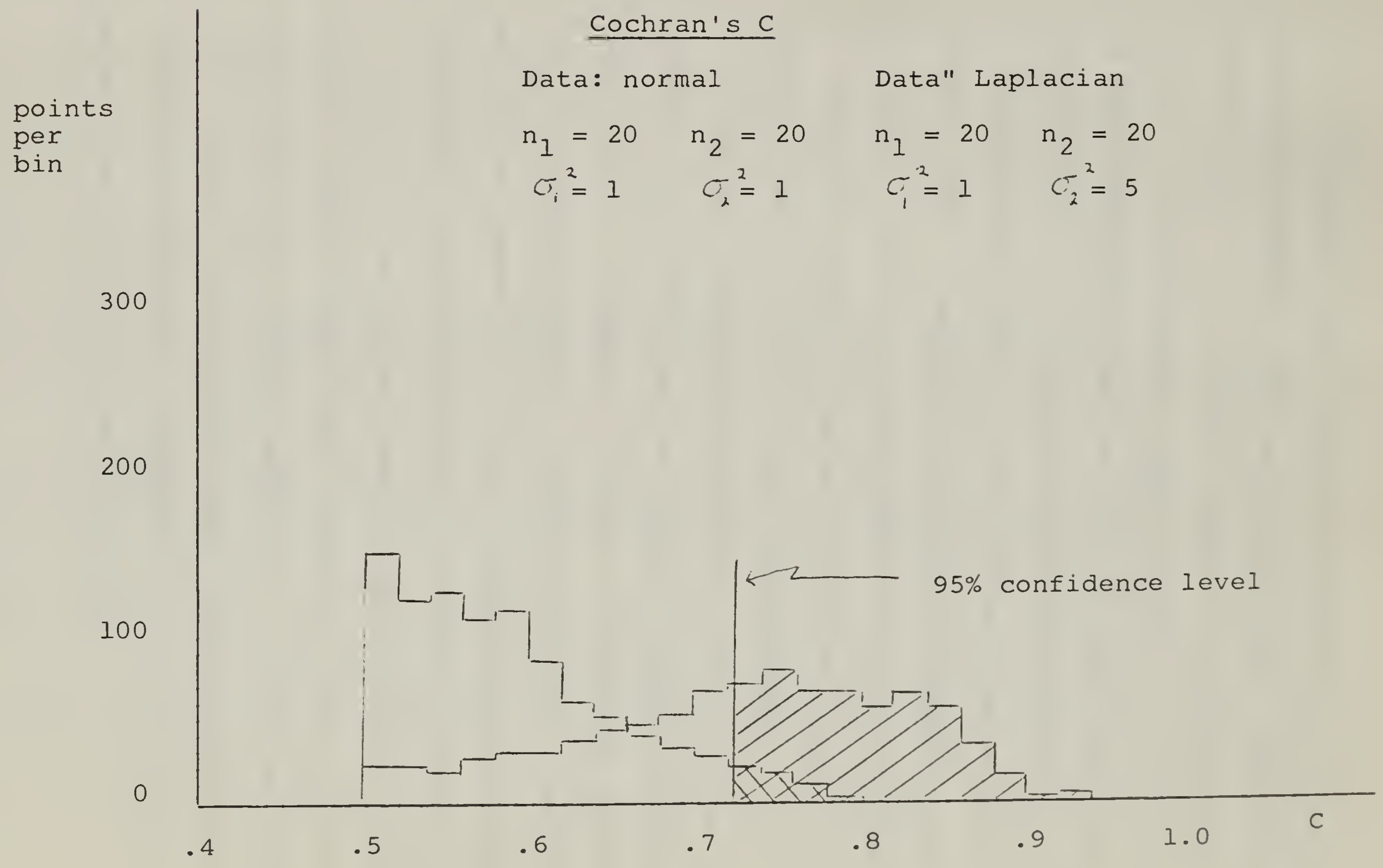

Cochran's C
Data: normal
$n_1 = 20$ $n_2 = 20$
$\sigma_1^2 = 1$ $\sigma_2^2 = 1$
Data" Laplacian
$n_1 = 20$ $n_2 = 20$
$\sigma_1^2 = 1$ $\sigma_2^2 = 5$
points per bin
300
200
100
0
95% confidence level
.4
.5
.6
.7
.8
.9
1.0
C

Figure 9

The tests. Both Cochran's and Hartley's tests are reasonably powerful under most conditions. High kurtosis weakens power, but skewness appears to counteract this effect. For the smallest sample sizes, $n_1 = 10$ $n_2 = 10$, the 99th percentile has low power but the 95th is considerably better. In cases such as this where the power is low, Hartley's test is a bit better than Cochran's. Miller's test is nearly as powerful as Cochran's and Hartley's. While it is less affected by low kurtosis than the other two tests, it is more affected by high kurtosis and in a few cases it has very little power.

The two versions of Scheffé's test are rather similar in their power with the five subcell version holding a slight edge. In general the test is much less powerful than the previous three. For most cases only the 90th percentile has respectable power. Furthermore the unequal sample size situation results in a lower power than might be expected. Here power seems to be a function of the smallest sample rather than (as was the case in the previous three tests) the average sample size. Once again high kurtosis results in the lowest power and while skewness does counteract this effect, it does so to a considerably lesser extent than it did for the other tests.

Monte Carlo Generated Power Values for Cochran's Tests

The alternative distribution $\sigma_2^2/\sigma_1^2 = 5$ is compared with the $\sigma_2^2/\sigma_1^2 = 1$ from normal data.

			$\gamma_1 = 0$			$\gamma_1 = 1.3$
N_1	N_2	$1-\alpha$	$\beta_2 = 1.8$	$\beta_2 = 3$	$\beta_2 = 6$	$\beta_2 = 5.4$
		.99	.204	.262	.116	.903
10	10	.95	.683	.590	.329	.982
		.90	.842	.721	.437	.994
		.99	.939	.828	.406	.999
20	20	.95	.984	.910	.565	1.000
		.90	.997	.964	.694	1.00
		.99	1.000	.976	.651	1.000
30	30	.95	1.000	.990	.791	1.000
		.90	1.000	.995	.836	1.000
		.99	.979	.917	.515	1.000
20	30	.95	.995	.965	.644	1.000
		.90	1.000	.988	.771	1.000
		.99	.601	.586	.297	.971
10	20	.95	.892	.799	.488	.988
		.95	.967	.891	.606	.991
		.99	.600	.587	.297	.975
10	30	.95	.957	.861	.557	.995
		.90	.994	.936	.663	.998

Table XI

Monte Carlo Generated Power Values for Hartley's Test

The alternative distribution $\sigma_2^2/\sigma_1^2 = 5$ is compared with the $\sigma_2^2/\sigma_1^2 = 1$ distribution from normal data.

			$\gamma_1 = 0$			$\gamma_1 = 1.3$
		$1-\alpha$	$\beta_2 = 1.8$	$\beta_2 = 3$	$\beta_2 = 6$	$\beta_2 = 5.4$
		.99	.274	.325	.143	.927
10	10	.95	.683	.590	.329	.982
		.90	.837	.713	.428	.994
		.99	.934	.820	.399	.999
20	20	.95	.984	.906	.561	1.000
		.90	.996	.961	.673	1.000
		.99	.999	.971	.627	1.000
30	30	.95	1.000	.989	.766	1.000
		.90	1.000	.994	.832	1.000
		.99	.982	.923	.523	1.000
20	30	.95	.992	.979	.682	1.000
		.90	1.000	.986	.754	1.000
		.99	.542	.547	.278	.966
10	20	.95	.887	.797	.483	.988
		.95	.964	.889	.599	.991
		.99	.698	.646	.343	.978
10	30	.95	.961	.867	.568	.995
		.90	.990	.929	.643	.998

Table XII

Monte Carlo Generated Power Values for Miller's Test

The alternative distribution $\sigma_2^2/\sigma_1^2 = 5$ is composed with the $\sigma_2^2/\sigma_1^2 = 1$ distribution from normal data.

N_1	N_2	$1-\alpha$	$\gamma_1 = 0$			$\gamma_1 = 1.3$
			$\beta_2 = 1.8$	$\beta_2 = 3$	$\beta_2 = 6$	$\beta_2 = 5.4$
		.99	.501	.248	.090	.679
10	10	.95	.896	.540	.234	.874
		.90	.914	.682	.346	.935
		.99	.972	.667	.204	.976
20	20	.95	.997	.890	.409	.994
		.90	.997	.940	.503	.998
		.99	1.000	.931	.340	.999
30	30	.95	1.000	.987	.599	1.000
		.90	1.000	.997	.727	1.000
		.99	.999	.853	.288	.997
20	30	.95	1.000	.941	.452	1.000
		.90	1.000	.972	.588	1.000
		.99	.636	.361	.152	.792
10	20	.95	.917	.644	.286	.960
		.95	.979	.800	.428	.991
		.99	.824	.489	.195	.924
10	30	.95	.966	.716	.337	.990
		.90	.987	.819	.454	.995

Table XIII

Monte Carlo Generated Power Values for Scheffe's Test

with two Subsamples

The alternative distribution $\sigma_2^2/\sigma_1^2 = 5$ is compared with the $\sigma_2^2/\sigma_1^2 = 1$ distribution from normal data.

			$\gamma_1 = 0$			$\gamma_2 = 1.3$
N_1	N_2	$1-\alpha$	$\beta_2 = 1.8$	$\beta_2 = 3$	$\beta_2 = 6$	$\beta_2 = 5.4$
		.99	.104	.046	.030	.119
10	10	.95	.347	.211	.119	.464
		.90	.553	.413	.224	.726
		.99	.255	.127	.048	.293
20	20	.95	.677	.419	.189	.765
		.90	.887	.665	.349	.938
		.99	.315	.190	.064	.333
30	30	.95	.883	.630	.308	.896
		.90	.987	.890	.504	.989
		.99	.319	.174	.079	.336
20	30	.95	.852	.629	.307	.902
		.90	.950	.818	.441	.971
		.99	.132	.070	.036	.134
10	20	.95	.454	.294	.183	.455
		.95	.768	.579	.348	.760
		.99	.178	.099	.064	.179
10	30	.95	.607	.416	.221	.594
		.90	.821	.622	.375	.798

Table XIV

Monte Carlo Generated Power Values for Scheffe's Test

with Two Subsamples

The alternative distribution $\sigma_2^2/\sigma_1^2 = 5$ is compared with the $\sigma_2^2/\sigma_1^2 = 1$ distribution from normal data

			$\gamma_1 = 0$			$\gamma_1 = 1.3$
N_1	N_2	$1-\alpha$	$\beta_2 = 1.8$	$\beta_2 = 3$	$\beta_2 = 6$	$\beta_1 = 5.4$
		.99	.085	.095	.045	.332
10	10	.95	.235	.256	.141	.575
		.90	.347	.373	.229	.709
		.99	.481	.295	.079	.799
20	20	.95	.753	.589	.293	.953
		.90	.880	.767	.454	.989
		.99	.885	.732	.279	.990
30	30	.95	.961	.914	.526	1.000
		.90	.985	.961	.683	1.000
		.99	.789	.604	.241	.866
20	30	.95	.943	.862	.542	.971
		.90	.976	.938	.694	.991
		.99	.261	.226	.132	.200
10	20	.95	.527	.455	.294	.356
		.95	.711	.655	.453	.481
		.99	.262	.202	.100	.125
10	30	.95	.587	.453	.300	.272
		.90	.844	.721	.540	.436

Table XV

Power curves. Ideally a discussion of power should include a power curve illustrating how power changes as a function of the alternative hypothesis. These curves are expensive to generate through a Monte Carlo technique like the one used in this study. Tables XVI and XVII present, for Cochran's and Hartley's tests, three points from each of a series of such curves. Column one is simply the confidence level since that is the power when the alternative hypothesis is no different from the null. Column two is the power for the alternative hypothesis $\sigma_1^2 = 1$ $\sigma_2^2 = 2$ and column three is the data given in tables XI and XII for the alternative $\sigma_1^2 = 1$ $\sigma_2^2 = 5$. These points are for the case of normal data only and do not reflect the robustness of the power function. In all situations substancial drops are found in the power rating when the alternative hypothesis changes from $\sigma_1^2 = 1$ $\sigma_2^2 = 5$ to $\sigma_1^2 = 1$ $\sigma_2^2 = 2$. Thus one cannot expect the performance of Hartley's and Cochran's tests to be excellent for any alternative hypothesis $\sigma_2^2/\sigma_1^2 < 5$. At the same time, since the power at $\sigma_2^2/\sigma_1^2 = 5$ is usually close to 1.0 little improvement can be expected for alternative distributions with $\sigma_2^2/\sigma_1^2 > 5$. Figure (10) is an attempt to guess at the shape of the power curve; it is based on the three points given in table XVI. It should be remembered, however, that several other curves will also fit the data.

Power Curves for Cochran's Test

Sample Sizes		σ_2^2/σ_1^2 1	2	5
10	10	.01	.026	.262
		.05	.136	.590
		.10	.236	.721
20	20	.01	.168	.829
		.05	.308	.910
		.10	.448	.964
30	30	.01	.331	.976
		.05	.529	.990
		.10	.640	.995
20	30	.01	.187	.917
		.05	.350	.965
		.10	.528	.988
10	20	.01	.090	.586
		.05	.231	.799
		.10	.346	.891
10	30	.01	.091	.587
		.05	.312	.861
		.10	.427	.936

Table XVI

Power Curves for Hartley's Test

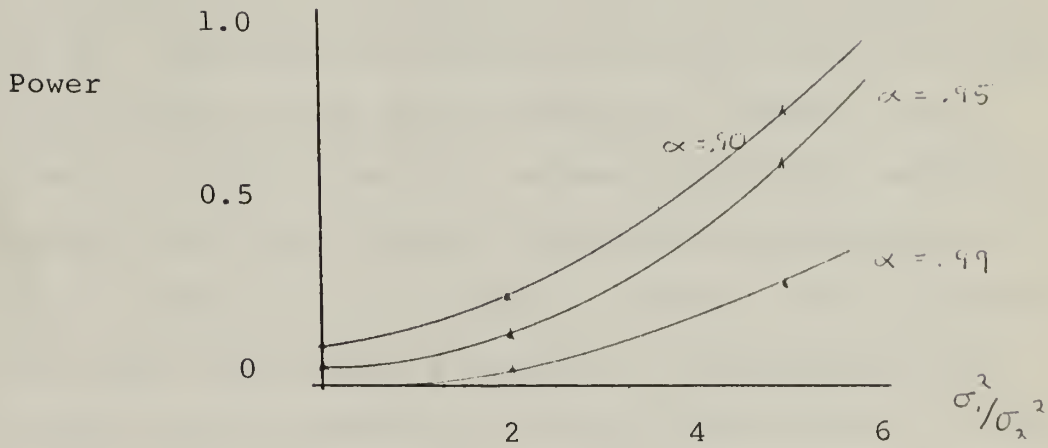
Sample Sizes		σ_2^2/σ_1^2 1	2	5
10	10	.01	.035	.325
		.05	.136	.590
		.10	.227	.713
20	20	.01	.165	.820
		.05	.299	.906
		.10	.422	.961
30	30	.01	.294	.971
		.05	.494	.989
		.10	.621	.994
20	30	.01	.172	.923
		.05	.361	.979
		.10	.495	.986
10	20	.01	.073	.547
		.05	.224	.797
		.10	.341	.889
10	30	.01	.109	.646
		.05	.322	.867
		.10	.409	.929

Table XVII

Power Curves for Cochran's Test

Three Significance Levels

$n_1 = 10 \quad n_2 = 10$



Three Sample Sizes

significance level = .01

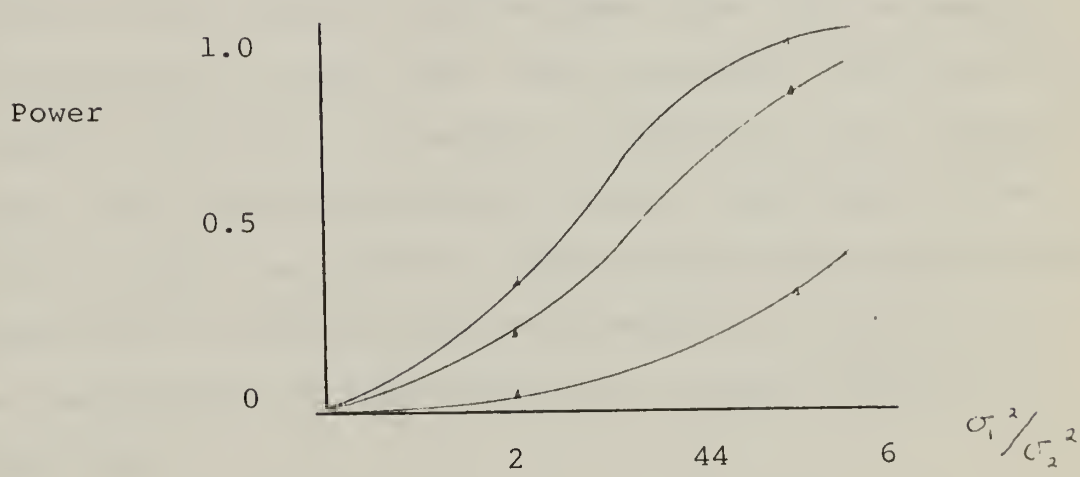


Figure 10

A Glance Behind the Tables

It is instructive to examine how shifts in the test statistic distributions cause the effects noted in the above tables. The only distributions that are displayed in this paper are for Cochran's test with both samples of size 20. These are the distributions from which the second rows of tables VI and XI were derived. Figures 1 to 4 present a simple situation; increasing kurtosis lengthens the right hand tail and thus the observed percentiles in table VI increase with kurtosis. Tables 5 to 8 illustrate a much more complicated situation. Increasing kurtosis flattens the distributions increasing both tails. In addition kurtosis values either above or below that for the normal (3) drive the distribution mean to the left. In figure 5 the short tail is more significant than the leftward drift of the mean and thus the power under the rectangular distribution is higher than under the normal (see table XI). Figure 8 shows that the effect of skewness is very significant. The flat distribution of figure 7 has been changed into a highly peaked short tailed distribution and the leftward drift of the mean has been completely reversed into a considerable rightward shift, as a result the power is very high.

A Note on Miller's Test

Miller's T test is a two tailed test; for this study the absolute value of Miller's T was used, creating a one tailed test. If the two tailed version is used a cautionary note about power is necessary. The Miller T distributions generated from data of unequal variance and rectangular, normal or Laplacian origin all fall to the left of the equal variance distribution but that generated from the doubly exponential distribution falls to the right. If the two tailed version is incorrectly used as a one tailed test then that test may have zero power and the probability of making an incorrect decision will be extremely high (probability of Type I error = 0 but probability of Type II error = 1).

CHAPTER V

CONCLUSIONS

A Warning

The results of this study provide strong evidence of the importance of knowing both the Type I and the Type II error probabilities. Scheffé's test is clearly superior if only Type I error is considered but it fails completely when Type II error is weighed.

The price that has been paid for the Type I robustness of the Scheffé test is that the power is uniformly low; too low for the test to be useful. Nor is this a surprising circumstance since it is often neccessary to sacrifice one test characteristic in order to improve another.

One of the more dangerous traditions of statistical researchers is to pick the most powerful test available without checking whether that test has adequate power for the particular application. The logic apparently is that since the test is the best available it is the one to use. However, if the power is sufficiently low then it may make far more sense to employ no test at all. In the case of preliminary tests, for example, accepting the null hypothesis permits one to proceed to the main test. It is therefore important that the null hypothesis not be accepted when it is false, i.e. that the Type II error probability be low. Consider the case of a test with power 0.1 at the 10% significance level. In this case both the null valid distribution and the alternative hypothesis distribution have nine tenths of their areas below the 90th percentile. If the test statistic falls below the 10% cut off and if the null hypothesis is accepted then there is a 50/50 chance that the correct decision has been made. (This assumes that there is no pretest information; i.e. the _a priori_ probability of H_o true is 1/2.) In brief a coin flip would have been as useful. If the power is still lower, then the test is biased against the correct decision and is worse than a coin flip. Nor is this a fictitious example: the power of Scheffe's test with $J_i = 2$, $n_1 = 10$, $n_2 = 10$ and data from a distribution of high kurtosis is 0.12 at the 10% significance level, which is just barely above the coin flip level.

Comparison of the Tests

Miller's test is better in overall performance than either Cochran's or Hartley's. However, this difference is slight compared to its vastly greater computational difficulty. Cochran's and Hartley's tests are very similar in performance and also in difficulty of computation. Cochran's test may be superior when sample sizes are unequal and when kurtosis is high; it is therefore preferable in spite of the slightly greater computation.

Thus based on the results of this study, and for the ranges of sample sizes considered, one can recommend Cochran's test for most situations with Miller's test a possibility for those cases in which a little more power is worth the price of much greater computation. (The power of the two tests is of course essentially the same but if Cochran's test is corrected for its lack of robustness in the confidence levels, then it loses some power.) For regions beyond the range of this study, in particular for larger n's or greater numbers of samples, or both, Scheffé's test may have sufficient power to become the preferable test. However, it is much more difficult to calculate than is Cochran's and it would hold an advantage only in those few cases where its greater Type I error robustness was of real value.

Two tests not considered in this study should also be mentioned, the Box - Andersen and the Bartlett. Miller (1968) studied the Box-Andersen test and found it very similar to the Miller Jackknife (see tables II & III). His study did not include unequal sample sizes so the performance under these conditions is unknown. Since Miller's test performs very well for unequal sample sizes it is unlikely that the Box-Andersen is better and it could be considerably worse. Thus until further information is gained it is safer, and probably at least as effective, to use the

Miller Jackknife rather than the Box-Andersen.

Bartlett's test was not considered in this study because it was believed to be sensitive to non-normality and it did not share Cochran's and Hartley's ease of calculation. However, the unexpectedly good performance of Cochran's and Hartley's tests suggests that Bartlett's may not be as poor as had been anticipated. It certainly should be investigated in future studies as a potential candidate for an improvement over Cochran's test at a reasonable calculational cost. My own guess is that it will actually prove to be very similar to Cochran's test in Type I error robustness and only slightly, if at all, more powerful.

Educational Implications

Surprisingly this study is not completely devoid of educational implications and applications. The Monte Carlo method of reproducing test statistic distributions gives a far more convincing demonstration of how statistical tests work than has ever been possible before. A student may design any experiment, formulate any series of conditions, consistent or inconsistent with the test assumptions, and then by use of a set of programs such as those included in the appendix he may literally watch the laws of chance construct a test statistic distribution. Next he may measure percentiles and power functions or observe how distributions change under the influence of various factors. Such manipulations give the student a far better grasp of the meaning of confidence level and power than can a few sample problems and several hours of flipping through a set of tables whose origin is obscure. Furthermore the concepts of test robustness and power take on a more intuitive meaning, at least for the tests he has investigated. Instead of viewing the field of statistical testing in terms of number tables whose values change in some random fashion he may form a dynamic picture of a distribution gradually changing shape and drifting either to the left or right along a test statistic axis. Figures 1 to 8 show for Cochran's test how the C distribution changes under the influence of non-normal data and violation of the null hypothesis. A brief study of these graphs is enough to generate an intuition as to how the curve would look for situations not considered. True similar results may be obtained from tables through interpolation or extrapolation, but these results lack some of the pictorial richness that begining students usually require.

The chief advantage of the Monte Carlo approach as a teaching tool is that it allows the student to compare rival tests in a pictorial manner and for a single test to compare situations in which it works to those in which it does not. As regions are reached in which the test's power becomes low it is graphically clear that the null distribution cannot be separated from the alternative hypothesis. Since the student is likely to encounter situations in which standard well known tests fail miserably in their performance, he will learn to exercise a greater degree of caution in his use of statistics than is presently the norm.

In addition to an increased awareness of the weakness of tests, the student may also gain an increased awareness of the weakness of numbers, or to use a more conventional phrase, significant figures. The malpractice of viewing experimentally derived numbers as absolute exact quantities whose value can be reported to any arbitrary accuracy has reached epidemic proportions in educational and psychological practice. There is some reason to hope that experience with Monte Carlo generated confidence levels can help to correct this problem by showing the student examples of the varying significance of numbers. One example is the error band that surounds each percentile, (see appendix II). A little familiarity with tables VI to X will convince one that the numbers reported there contain more digits than is really useful. In the first entry of table VI (0.794) the 4 contains no useful information at all and mearly serves to confuse. Tables XI to XV do not give error bars and may easily mislead the reader into believing in the significance of numbers that differ only in their third decimal place.

Related to the problem of significant figures is the effect that thin tails have on the significance of exact percentiles. In Scheffé's test where the tail is very long and thin, changing the cut off by 10% makes virtually no difference to the α level; but in Cochran's test a 10% change in the cut off changes the confidence level considerably. Thus while it is useful to know the Cochran cut offs to three significant figures, two are more than adequate for Scheffé's test.

BIBLIOGRAPHY

Bartlett, M.S. and Kendall, D.G., The statistical analysis of variance heterogeneity and the logarithmic transformation. Journal of the Royal Statistical Society, 1946, sup. 8 (1), 128-138.

Bartlett, M.S., Properties of sufficiency and statistical tests. Proceedings of the Royal Society of London, 1937, A 160, 268-282.

Box, G.E.P., Non-normality and tests on variances. Biometrika, 1953, 40, 318-335.

Box, G.E.P., Some Theorems on Quadratic Forms applied in the study of analysis of variance problems, 1. Effect of inequality of variance in the one-way classification. Annals of Mathematical Statistics, 1954, 25, 290-302.

Box, G.E.P., Some Theorems on Quadratic forms applied in the study of analysis of variance problems, II, Effects of inequality of variance and of correlation between errors in the two-way classification. Annals of Mathematical Statistics, 1954, 25, 484-498.

Box, G.E.P. and Andersen, S.L., Permutation Theory in the derivation of robustness criteria and the study of departures from assumption. Journal of the Royal Statistical Society, 1955, Series B, Vol. XVII #1, 1-34.

Box and Muller, A note on the generation of normal deviates. Annals of Mathematical Statistics, 1958, 28, p. 610-611

Bradley, J.V., Studies in Research Methodology IV. AMRL Technical Document Report, 1963, 63-69, Wright-Patterson Air Force Base, Ohio.

Bradley, J.V., Studies in Research Methodology VI. AMRL Technical Document Report, 1964, 64-123, Wright-Patterson Air Force Base, Ohio.

Cochran, W.G., The distribution of the largest of a set of estimated variances as a fraction of their total. Annals of Eugenics, 1941,11, 47-52.

Cochran, W.G., Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics, 1947, 3, #1, 22-53.

David, F.N. and Johnson, N.L., The effect of non-normality on the power function of the F-test in the analysis of variance. Biometrika, 1951, 38, 43-57.

Donaldson, T.S., Robustness of the F-test to errors of both kinds and the correlation between the numerator and denominator of the F-ratio. Journal of the American Statistical Association, 1968, 660-676.

Durand, A.L., Comparitive power of various tests of homogeneity of variance. Master's Thesis, School of Education, University of Colorado, 1959.

Eisenhart, C., The assumptions underlying the analysis of variance. Biometriks, 1947, 3 #1, 1-21.

Glass, G.V., Testing homogeneity of variances. American Education Research Journal, 1966, 3 (3), 187-190.

Gronow, D.G.C., Test for the significance of the difference between means in two normal populations having unequal variances. Biometrika, 1951, 38, 252-256.

The Handbook of Mathematical Functions. Edited by Abromowitz, M. and Stegum, A., National Bureau of Standards, 1964.

Hartley, H.O., The maximum F-ratio as a short-cut test for heterogeneity of varience. Biometrika, 1950, 37, 308-312.

Hsu, P.L., Contribution to the theory of "students" t-test as applied to the problem of two samples. Statistical Research Memoirs, 1938, Vol. II, 1-24.

Hsu, P.L., Analysis of variance from the power function standpoint. Biometrika, 1941, 32, 62-69.

Levene, H., Robust tests for equality of variances. Contributions to Probability and Statistics, Edited by Olkin, I., Standford Univ. Press, 1960.

Lindquist, E.F., Design and Analysis of Experiments in Psychology and Education, Houghton Mifflin Co., Boston, 1953, 78-90.

Miller, R.G. Jr., Jackknifing variances. The Annals of Mathematical Statistics, 1968, 39 (2), 567-582.

Moses, L.E., Rank tests of dispersion. Annals of Mathematical Statistics, 1963, 34, 973-983.

Myers, J.L., private communication.

Odeh, R. and Olds, E.G., Notes on the analysis of variance of logarithms of variances, Springfield, Va. Clearing for federal scientific and technical information. March, 1959, PB 161 108.

Scheffe, H., The Analysis of Variance, New York, Wiley, 1959.

Tang, P.C., The power function of the analysis of variance tests with tables and illustrations of their use, Statistical Research Memoirs, 1938, Vol. II, 126-157.

Tocher, K.D., The art of simulation, The English Universities Press Ltd., London, 1963.

Wald, A., Sequential Analysis, New York, Wiley, 1947.

Welch, B.L., The significance of the difference between two means when the population variances are unequal. Biometrika, 1938, 29, 350.

Appendix I

Skewness (γ_1) and kurtosis (β_2) are related in such a way that the value of β_2 limits the value of γ_1. In particular for the case mean zero ($\mu_1 = 0$) one can show that

$$\beta_2 \geq 1 + \gamma_1^2$$

<u>Proof</u>. Let x be a random variable with $\mu_1(x) = 0$ and let $y = x^2$. From the Cauchy Swartz inequality we know that

$$\sigma_{xy}^2 \leq \sigma_x^2 \sigma_y^2$$

where: $$\sigma_{xy} = E(xy) + E(x) E(y) = E(x^3) = \mu_3$$

$$\sigma_y^2 = E(y^2) - (E(y))^2$$

$$= E(x^4) - (E(x^2))^2 = \mu_4 - \mu_2^2$$

Thus: $$\mu_3^2 \leq \mu_2(\mu_4 - \mu_2^2)$$

Subtituting: $$\gamma_1^2 = \mu_3^2/\mu_2^3 \text{ and } \beta_2 = \mu_4/\mu_2^2$$

we get: $$\gamma_1^2 = \beta_2 - 1$$

Appendix II

The Confidence band about the Confidence Level

There are two major sources of error associated with each confidence level. One derives from the nature of the random number generator and the approximate techniques used to simulate the various distributions. This error has been ignored because 1) it is believed to be small and 2) it is not significant to this study since it has the effect of only shifting the moments in the parent distribution slightly. The second source of error stems from the finite size of the statistic distribution, one thousand points, and is readily estimated.

Two techniques for calculating percentile confidence intervals are presented. The first is based on the theory of order statistics and may be beyond the mathematical sophistication of many readers. The second is an intuitive argument that avoids the complications of the first technique by solving a different problem and then drawing an analogy. I am indebted to Robert Kleyle for showing me the first argument and to Richard Kofler for explaining the second.

Exact Estimation of the Confidence Intervals for Percentiles

Let $F(x)$ be the cumlative frequency distribution (c.d.f.) of some arbitrary distribution $f(x)$. Define E_p as the p^{th} percentile where $P(x \leq E_p) = F(E_p) = p$.

Let $x_1, x_2, \ldots, x_n$ denote a random sample from a continuous distribution $F(x)$. Then if $Y_i = F(x_i)$, $i = 1, \ldots, n$, $Y_1, \ldots, Y_n$ are independent and uniformly distributed on $(0,1)$. Furthermore, since c.d.f.'s are non-decreasing, $Y_{(1)} \leq Y_{(2)} \leq \ldots, \leq Y_{(n)}$ are the order statistics of a random sample (size n) from the uniform distribution. Thus the joint pdf of $(Y_{(1)}, \ldots, Y_{(n)})$ is

$$g(y_1, \ldots, y_n) = n! \qquad 0 < y_1 < y_2 < \ldots < y_n < 1$$
$$0 \qquad \text{elsewhere}$$

$\therefore$ The joint pdf of $Y_{(k_1)}, Y_{(k_2)}$ $k_1 < k_2$ is $g_{k_1 k_2}(y_1, y_2)$

$$= \frac{(n!)\left[y_1^{k_1-1}(y_2-y_1)^{k_2-k_1-1}(1-y_2)^{n-k_2}\right]}{(k_1-1)!(k_2-k_1-1)!(n-k_2)!} \qquad 0<y_1<y_2<1$$

$$= 0 \qquad \text{elsewhere}$$

(See Hogg & Graig pp. 179-180)

Now notice that since F is strictly increasing,

$$x_{(k_1)} < E_p < x_{(k_2)} \Longleftrightarrow$$

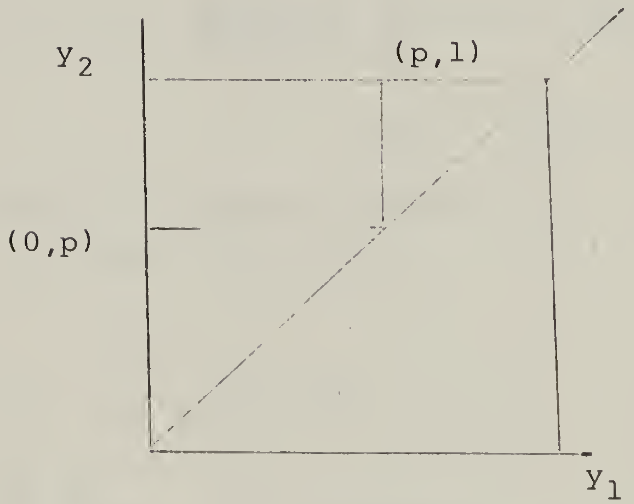
$$F(x_{(k_1)}) < p < F(x_{(k_2)}).$$

$$\therefore P(x_{(k_1)} < E_p < x_{(k_2)}) = P(Y_{(k_1)} < p < Y_{(k_2)})$$

$$= \int_p^1 \int_0^p g_{k_1k_2}(y_1 y_2)dy_1 dy_2$$

$$= \int_0^p \int_{y_1}^1 g_{k_1k_2}(y_1, y_2)dy_2 dy_1$$

$$- \int_0^p \int_0^{y_2} g_{k_1k_2}(y_1, y_2)dy_1 dy_2$$



To evaluate the first integral,

let $y_1 = u, \quad y_2 = 1 - v(1 - u)$

$$J = \begin{vmatrix} 1 & 0 \\ v & -(1-u) \end{vmatrix} = -(1 - u)$$

$$|J| = 1 - u$$

$$\therefore \int_0^p \int_{y_1}^1 g_{k_1k_2}(y_1,y_2)dy_2dy_1$$

$$= \frac{n!}{(k_1-1)!(k_2-k_1-1)!(n-k_2)!} \int_0^p u^{k_1-1}(1-u)^{n-k_1} \int_0^1 v^{n-k_2}(1-v)^{k_2-k_1-1} dvdu$$

$$= \frac{n!}{(k_1-1)!(n-k_1)!} \int_0^p u^{k_1-1}(1-u)^{n-k_1}du$$

$$= Ip(k_1, n-k_1 + 1)$$

where $I_\theta(a,b)$ denotes the incomplete beta function,

$$I_\theta(a,b) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \int_0^\theta x^{a-1}(1-x)^{b-1}dx$$

$$a,b>0 \qquad \theta \in (0,1).$$

To evaluate the second integral let

$$y_1 = uv, \qquad y_2 = v.$$

$$J = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v>0$$

$$\therefore \int_0^p \int_0^{y_1} g_{k_1k_2}(y_1,y_2)dy_1dy_2$$

$$= \frac{n!}{(k_1-1)!(k_2-k_1-1)!(n-k_2)!} \int_0^p v^{k_2-1}(1-v)^{n-k_2}$$

$$\int_0^1 u^{k_1-1}(1-u)^{k_2-k_1-1}dudv$$

$$= \frac{n!}{(k_2-1)!(n-k_2)!} \int_0^p v^{k_2-1}(1-v)^{n-k_2}dv$$

$$= Ip(k_2, n-k_2 + 1).$$

$$\therefore \quad P(x_{(k_1)} < Ep < x_{(k_2)})$$

$$= Ip(k_1, n-k_1 + 1) - Ip(k_2, n-k_2 + 1),$$

and $(x_{(k_1)}, x_{(k_2)})$ forms a confidence interval for Ep with confidence coefficient

$$\gamma(k_1, k_2) = Ip(k_1, n - k_1 + 1) - Ip(k_2, n - k_2 + 1).$$

Asymptotic Intervals

Recall that

$$P(x_{(k_1)} < Ep < x_{(k_2)}) = Ip(k_1, n-k_1+1) - Ip(k_2, n-k_2+1)$$

$$= \sum_{x=k_1}^{k_2-1} \binom{n}{x} p^x (1-p)^{n-x}$$

Thus if T_n has the binominal (n, p) distribution,

$$P(x_{(k_1)} < Ep < x_{(k_2)}) = P(k_1 \le T_n \le k_2-1 \mid p).$$

But since

$$\frac{T_n - np}{\sqrt{np(1-p)}} \xrightarrow[d]{} N(0, 1)$$

it is clear that for sufficiently large

$$P(k_1 \leq T_n \leq k_2-1) \simeq P(k_1-1/2 < u_n < k_2-1/2)$$

where u_n has the normal distribution with parameters $\mu = np$ and $\sigma^2 = np(1-p)$.

Note: The 1/2 added to $k_2 - 1$ and subtracted for k_1 is the correction for continuity. i.e. the correction used when approximating a discrete distribution with a continuous distribution. ∴ For large n, if Φ denotes the c.d.f. of N(0, 1)

$$P(x_{(k_1)} < Ep < x_{(k_2)})$$

$$\sim \Phi\left(\frac{k - 1/2 - np}{\sqrt{np(1-p)}}\right) - \left(\frac{k_1 - 1/2 - np}{\sqrt{np(1-p)}}\right)$$

We will now calculate the confidence coefficient for the three percentiles used in this study using predetermined values for k_1 and k_2. For the origin of these values see the next section. In all cases n = 1000. For the values of $\Phi(t)$ refer to table XVIII.

The 99th percentile. $p = .99 \quad k_1 = 987 \quad k_2 = 993$

$P(x_{987} < \xi_p < x_{993}) =$

$$\Phi\left(\frac{993 - 1/2 - 990}{\sqrt{990 \times .01}}\right) - \Phi\left(\frac{987 - 1/2 - 990}{\sqrt{990 \times .01}}\right)$$

$$= \Phi\left(\frac{2.50}{3.15}\right) - \Phi\left(\frac{-3.50}{3.15}\right)$$

$$= \Phi(.79) - \Phi(-1.11) = .7852 - .1335$$

$$= .65$$

The 95th percentile $p = .95 \quad k_1 = 943 \quad k_2 = 957$

$P(x_{943} < \xi_p < x_{957}) =$

$$\Phi\left(\frac{957 - 1/2 - 950}{\sqrt{950 \times .05}}\right) - \Phi\left(\frac{943 - 1/2 - 950}{\sqrt{950 \times .05}}\right)$$

$$= \Phi\left(\frac{6.50}{6.89}\right) - \Phi\left(\frac{-7.50}{6.89}\right) = \Phi(.94) - \Phi(-1.09)$$

$$= .69$$

The 90th percentile $p = .90 \quad k_1 = 890 \quad k_2 = 910$

$P(x_{890} < \xi_p < x_{910}) =$

$$\phi\left(\frac{910 - 1/2 - 900}{\sqrt{900 \times .1}}\right) - \phi\left(\frac{890 - 1/2 - 900}{\sqrt{900 \times .1}}\right)$$

$$= \phi\left(\frac{9.50}{9.49}\right) - \phi\left(\frac{-10.50}{9.49}\right)$$

$$= \phi(1.00) - \phi(-1.11)$$

$$= .71$$

An Intuitive Estimate of the Confidence Interval for Percentiles

Let x be the number of points in the bin falling to the right of the percentile and n - x the number of points falling to the left. x is distributed as a binomial B(n, p) where p is the probability of an event occuring in the right bin. The central limit theorem states that the x distribution can be approximated by a normal $N(\mu = np, \sigma = np(1-p))$. This approximation is good when ever the minimum of np and $n(1-p) \gtrsim 5$ which for n = 1000 and $p \geq .01$ is the case. Thus

$$P\left(-1 < \frac{x-np}{\sqrt{np(1-p)}} < 1\right) \simeq .68$$

or

$$P\left(np - \sqrt{np(1-p)} < x < np + \sqrt{np(1-p)}\right) \simeq .68$$

np may be replaced by its estimator x. Since $x << n$

$$\sqrt{\frac{nx - x^2}{n}} \approx \sqrt{x}$$

Thus if x points are found in a bin there is a 68% probability that the true value of np falls within an interval $x \pm \sqrt{x}$.

When dealing with experimentally generated confidence levels one is faced with a slightly different problem than that just solved. Rather than defining the limits of a bin and then counting the number of points, a number of points is counted to define the lower limit of a bin; the bin being that bin that contains $(1-\alpha)\%$ of the total number of points and is bounded by ∞ on the upper end and the α cut off on the lower. Suppose that the true value of the confidence level were known and several finite distributions were generated to test this level. One would then expect that 68% of the time the experiment would find $(1-\alpha)N \pm \sqrt{(1-\alpha)N}$ points above the confidence level. This information can be turned around to say that if one finds the points associated with bins of size $(1-\alpha)N + \sqrt{(1-\alpha)N}$ and of size $(1-\alpha)N - \sqrt{(1-\alpha)N}$ then one has made a best estimate of a region within which the α confidence level lies with a probability of 68%.

The values given in tables VI to X are the lower limits for the bins containing $(1 - \alpha)N$ points, i.e. the α cut offs. To the right and above each such number is the lower limit for the bin containing $(1 - \alpha)N + \sqrt{(1 - \alpha)N}$

points and below that the lower limit of the bin containing $(1-\alpha)N - \sqrt{(1-\alpha)N}$ points. Between these two limits is the region in which the true α level cut off lies with probability $\sim$ 68%.

It may now be noted that the values for k_1 and k_2 given in the previous section are $(1-\alpha)N \pm \sqrt{(1-\alpha)N}$.

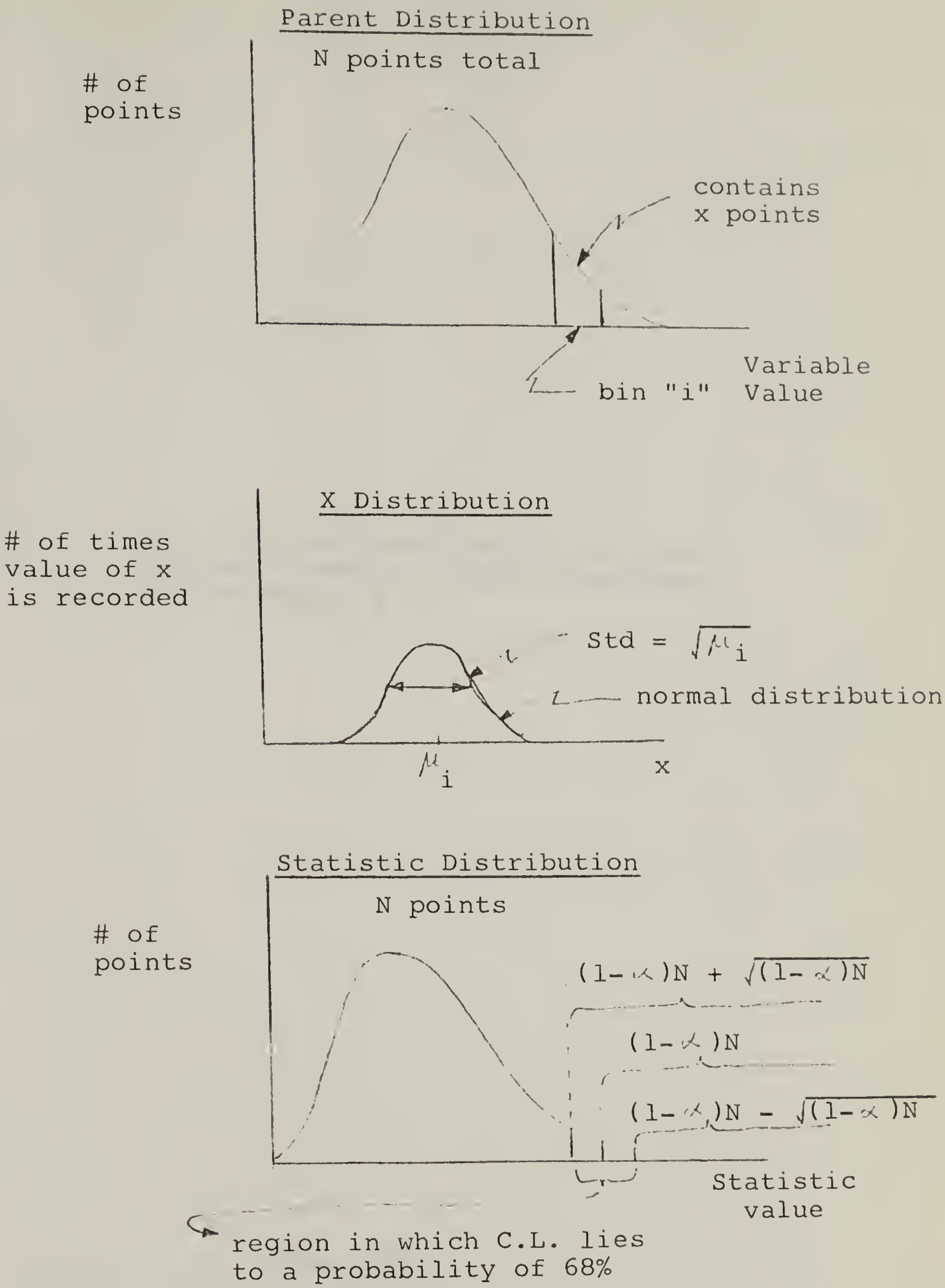

Parent Distribution
N points total
# of points
contains x points
bin "i"
Variable Value
X Distribution
# of times value of x is recorded
Std = $\sqrt{\mu_i}$
normal distribution
μ_i
x
Statistic Distribution
N points
# of points
$(1-\alpha)N + \sqrt{(1-\alpha)N}$
$(1-\alpha)N$
$(1-\alpha)N - \sqrt{(1-\alpha)N}$
Statistic value
region in which C.L. lies to a probability of 68%

Figure 11

(see any basic statistics text for a table of the area under the Normal distribution)

Table XVIII

Appendix III

The Tests

Cochran's Test. Cochran's test is described in Winer (1962) p. 94.

For J samples with variance s_j^2 and all size n.

$$C = \frac{s^2 \text{ largest}}{\sum_{j=1}^{J} s_j^2}$$

A table of C values appears in Winer p. 654, parametrized by df. s^2 and J. All samples are supposed to be of equal size and the df. associated with s^2 is n-1.

Hartley's Test. Hartley's test is described in Winer (1962) p. 93 - 94.

$$F_{max} = \frac{\text{the largest of the cell variances}}{\text{the smallest of the cell variances}}$$

A table of F_{max} appears in Winer p. 653, parametrized by df. s^2 and the number of cells. Each cell is of size n and the df. for s^2 is n-1.

Miller's Test. Miller's test for the two group comparison is described in Miller (1968). Two samples x and y of size N and M are divided into subsamples of size k (k = 1 gives the most power).

$$n = N/k \qquad m = M/k$$

s_{x-i}^2 is the variance of group x with the i^{th} subgroup deleted.

s_x^2 is the variance of group x

σ_x^2 is the variance of the population from which group x is drawn.

$\theta_x = \text{Log } \sigma_x^2$

$\bar{\theta}_x = \text{Log } s_x^2$

$\bar{\theta}_{xi} = n \log s_x^2 - (n-1) \log s_{x-i}^2$

$\bar{\theta}_{x.} = n \log s_x^2 - \frac{(n-1)}{n} \sum_{i=1}^{n} \log s_{x-i}^2$

Then:

$$\frac{(\bar{\theta}_{x.} - \theta_x) - (\bar{\theta}_{y.} - \theta_y)}{\sqrt{\frac{1}{n(n-1)} \cdot \sum_{i=1}^{n} (\bar{\theta}_{xi} - \bar{\theta}_{x.})^2 - \frac{1}{m(m-1)} \cdot \sum_{i=1}^{m} (\bar{\theta}_{yi} - \bar{\theta}_{y.})^2}}$$

is distributed as t with n + m - 2 degrees of freedom. For the situation considered in this paper σ^2 can be set equal to 1 and $\theta_x = \theta_y = 0$.

<u>Scheffé's Test</u>. Scheffe's test is described in Scheffé (1959) p. 83-87.

Given I samples the test is for the hypothesis $\sigma_1^2 = \sigma_2^2 = \ldots \sigma_I^2$. Divide each sample into J_i subsamples. Each subsample is of size n_{ij} and has variance s_{ij}^2

Define:

$$y_{ij} = \log s_{ij}^2$$

$$n_i = \sum_j n_{ij}$$

$$\nu_{ij} = n_{ij} - 1 \qquad \nu_i = \sum_j \nu_{ij}$$

$$\nu_e = \sum_i (J_i - 1) \qquad \nu = \sum_i \nu_i$$

$$\eta_i = \frac{\sum_j n_{ij} y_{ij}}{n_i} \qquad \eta = \frac{\sum_i \nu_i \eta_i}{\nu}$$

Then

$$\frac{\nu_e}{I-1} = \frac{\sum_i \nu_i (\hat{\eta}_i - \eta)^2}{\sum_i \sum_j \nu_{ij} (Y_{ij} - \hat{\eta}_i)^2}$$

is distributed as F with I - 1 df. numerator and ν_e df. denominator.

Box-Andersen Test. The Box-Andersen test (for the two group comparison) is described in Box and Andersen (1955). It is a standard F ratio

$$F = \frac{n_2 \sum_{i=1}^{n_1} x_{1i}^2}{n_1 \sum_{j=1}^{n_2} x_{2j}^2}$$

but the degrees of freedom are dn_1 and dn_2 instead of n_1-1 and n_2-1

$$d = \left[1 + 1/2 \left(\frac{N+2}{N-1-(b_2-3)} \right) (b_2-3) \right]^{-1}$$

where $N = n_1 + n_2$

and

$$b_2 = \frac{(N+2) \left(\sum_{i=1}^{n_1} x_{1i}^4 + \sum_{j=1}^{n_2} x_{2j}^4 \right)}{\left(\sum_{j=1}^{n_1} x_{1i}^2 + \sum_{j=1}^{n_2} x_{2j}^2 \right)^2}$$

For given sample sizes, n_1 and n_2, there are a large number of possible degrees of freedom, the proper set being determined by d. Thus there is no unique F distribution for any given n_1 and n_2.

Appendix IV

Definitions of Moments and Coefficients

Mean: $m = \int_{-\infty}^{\infty} x f(x) dx$

n^{th} central moment:

$$\mu_n = \int_{-\infty}^{\infty} (x - m)^n f(x) dx$$

Variance:

$$\sigma^2 = \mu_2 = \int_{-\infty}^{\infty} (x - m)^2 f(x) dx$$

Skewness: $\gamma_1 = \mu_3 / \sigma^3$ or $\beta_1 =$

Kurtosis: $\beta_2 = \mu_4 / \sigma^4$ or $\gamma_2 =$

Appendix V

This appendix includes listings of the programs used in the study and instructions on their use. The programs were deliberately written to make it impossible for the totally nieve user to run them. The philosophy behind this perversness was developed from observation of the miss-use of packaged programs such as the BMD series, where it is possible for a user to enter data and receive results without having any understanding of the programs. Frequently the result of this simplicity is that the user is not aware of program malfunctions or that the program may be solving a different problem from the one intended. To avoid these errors it is necessary for the user to have at least a partial understanding of any program he uses. The instructions given for the enclosed programs assume that the user will read and study the programs themselves as well as the instruction sheets.

I Program Name

Program DISTEST

II Purpose

To test distributions generation programs, plot the distributions and calculate their mean, variance, skewness and kurtosis.

III How to use

The distribution program is added as a subroutine; the call statement (CALL RECTANG (x)) is changed to fit the subroutine.

1) Dictionary of terms

A - the number of points in the distribution.

```
      PROGRAM DISTEST
      DIMENSION DATA(10000)
      COMMON/HISTGRAM/NHIST,XMIN,XMAX,DELX,NBINS
      TYPE REAL KURT
      A=1000.0
      KA=A
      T=TIMEF(T)
      CALL RANFSET(T)
      NHIST=1
      XMIN=-4.0
      XMAX=4.0
      DELX=0.1
      NBINS=0
      CALL SETLIM(X,1.)
      DO 10 J=1,KA
      CALL RECTANG(X)
      NHIST=1
      CALL HISTSUMS(X)
   10 DATA(J)=X
      XTOT=0.0
      DO 20 J=1,KA
   20 XTOT=XTOT+DATA(J)
      XM=XTOT/A
      Z4=0.0
      Z3=0.0
      Z2=0.0
```

```
      DO 30 J=1,KA
      Z2=(DATA(J)-XM)**2+Z2
      Z3=(DATA(J)-XM)**3+Z3
   30 Z4=(DATA(J)-XM)**4+Z4
      VAR=Z2/(A-1.0)
      STD=SQRT(Z2/(A-1.0))
      KURT=Z4/(A*(STD**4))
      SKEW=Z3/(A*(STD**3))
      PRINT 90
   90 FORMAT(25H    MEAN    VAR   SKEW   KURT)
      PRINT 100,XM,VAR,SKEW,KURT
  100 FORMAT(1X,4F6.3)
      NHIST=1
      CALL HISTPLOT
      END
```

I Program Name

Program ROBUST

II Purpose'

To run Monte Carlo simulations of test statistic distributions. Several test statistics may be investigated at the same time.

III How to use.

Internal statement changes are required during set up. The program can only be run by a user who has read and understood the flow of the program. Refer to section three of Chapter III.

1) Dictionary of common variables

NRUNS - number of program runs

CL - confidence level or percentile of normal test distribution

NS - number of samples

NSAMP - sample size

NSCHRUN - number of Scheffé tests to be performed

NNK - number of subcells in each Scheffé test

Z, tt - dummy variables

K - number of points in test statistic distribution

2) Choice of configuration

Two listings for Program ROBUST are included to illustrate the wide variety of configurations available; they are refered to as P_1 and P_2.

a) If samples are all from the same distribution use set up P_1.

b) If samples are from different distributions use set up P_2. The statements just prior to #12 determine how samples are selected.

c) If power is to be calculated the C.L. array must be read in, see P_2. A call to POWER is also added.

d) If only one test statistic is to be investigated use set up P_1.

e) If several test statistics are to be investigated refer to P_2. CALL statements are added after statement 30. The plot loop after #40 may need modification. Additional calls to Histgram must be set up after #5.

3) Data cards are used to determine the number of samples to be used (2 for all cases in this study) and the number of observations in each sample. They also determine the number of subsamples used in Scheffé's test. When power is calculated the percentiles against which the power is measured are read in on data cards.

DATA CARDS

COLUMN		FORMAT
1	number of PROGRAM RUNS (≤ 9)	I1
1-50	CONFIDENCE levels for null hypothesis	5(F8.3,2x)
1-50	"	
1-50	" USE IN POWER	
1-50	" VERSION ONLY	
1-50	"	
1-50	"	
1	number of samples (≤ 5)	I1
1-10	size of each sample	I2
1	number of Scheffé Test	I1
1-10	k value for each sample	I2
⋮	repeat one card for each Scheffé test	

repeat all except 1st card one set for each program run.

Sample Data Cards for a Type I Error Run

6

2

1010

2

0202

0505

2

2020

2

0202

0505

2

3030

2

0202

0505

2

2030

2

0202

0505

2

1020

2

0202

0505

2

1030

2

0202

0505

Sample Data Cards for a Type II Error Run

6

.880	6.70	2.97	100.	9.25
.907	11.40	3.3	80.	10.3
.803	4.08	2.03	20.	4.71
.854	6.96	2.44	20.43	5.4
.763	3.22	1.65	8.80	3.10
.818	4.86	1.9	10.14	3.43

2

1010

2

0202

0505

.768	3.30	2.85	83.	12.60
.85	5.6	2.85	100.	10.8
.725	2.63	2.08	19.0	6.10
.792	3.68	2.12	22.	5.4
.691	2.22	1.76	8.46	3.68
.752	3.04	1.75	7.87	3.72

2

2020

2

0202

0505

.717	2.5	2.72	100.	9.6
.797	3.9	3.2	65.	11.3
.671	2.05	1.98	18.0	5.17

.743	2.88	2.17	15.	4.8
.645	1.82	1.58	7.82	3.47
.706	2.42	1.84	8.09	3.23

2

3030

2

0202

0505

.754	2.94	2.63	82.	10.6
.818	4.5	2.87	100.	11.4
.703	2.36	2.03	14.	5.30
.763	3.2	2.21	23.67	5.77
.669	2.02	1.70	7.25	3.31
.726	2.64	1.83	9.56	3.65

2

2030

2

0202

0505

.830	4.95	3.10	100.	13.40
.908	8.8	3.6	100.	14.
.769	3.28	2.15	21.29	8.0
.84	4.68	2.36	19.2	7.5
.728	2.70	1.68	8.89	5.47
.801	3.67	1.99	7.29	4.91

2

1020

2

0202

0505

.820	4.40	2.80	87.	17.17
.88	7.2	3.5	100.	15.
.746	2.90	2.08	17.	9.60
.814	4.37	2.39	24.	10.6
.711	2.45	1.66	8.30	5.91
.776	3.49	1.91	12.25	7.0

2

1030

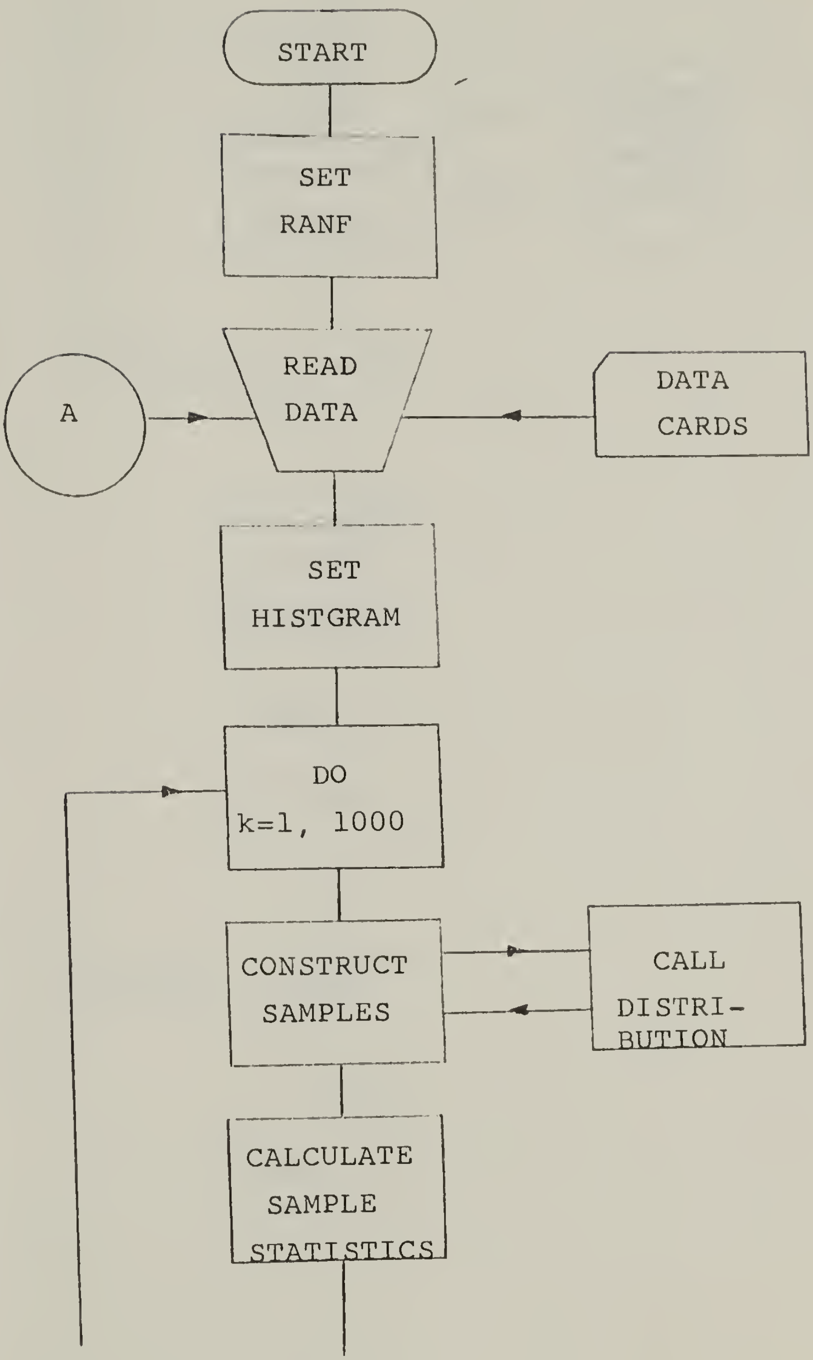
2

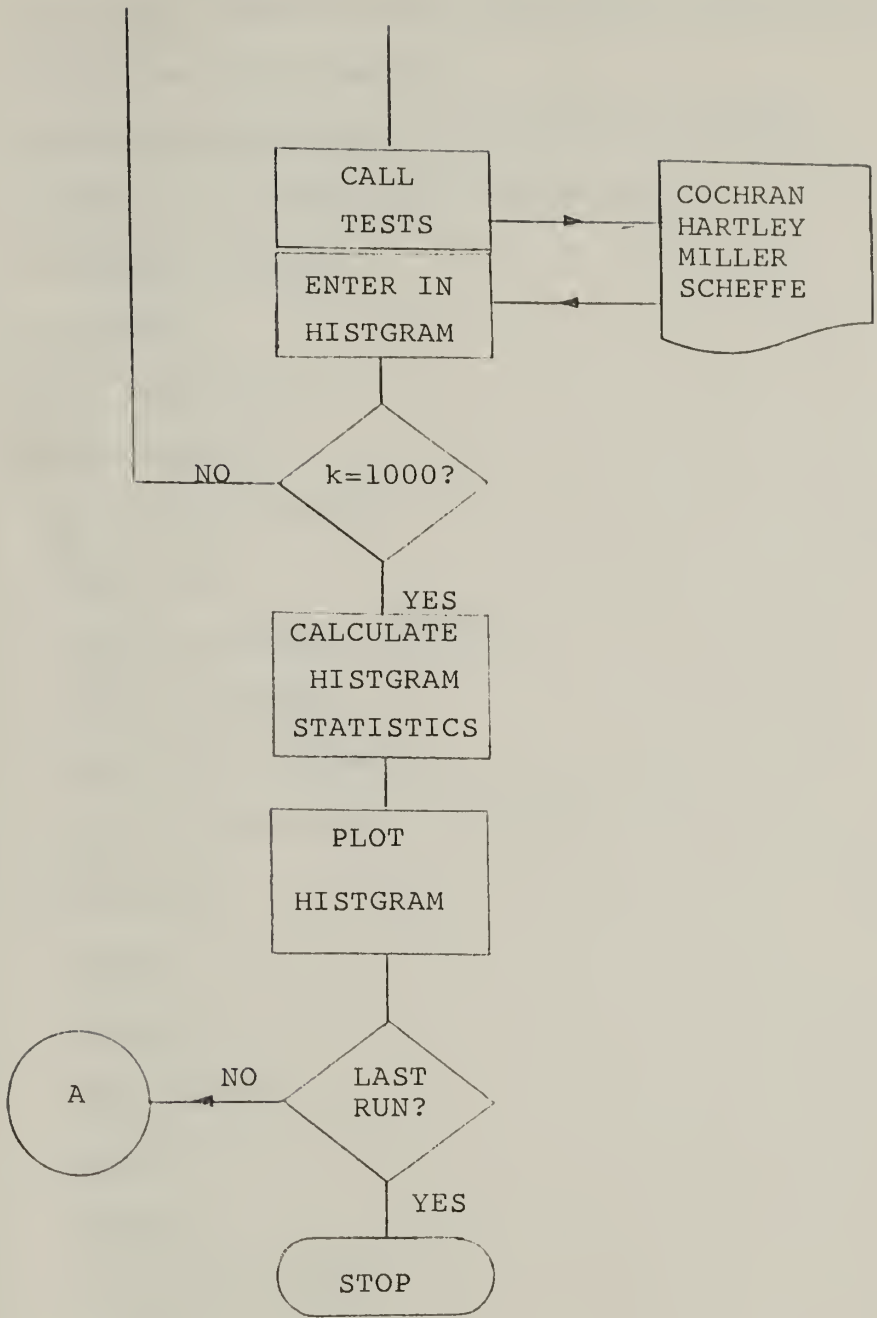
0202

0505

@@

PROGRAM ROBUST




CALL
TESTS
ENTER IN
HISTGRAM
COCHRAN
HARTLEY
MILLER
SCHEFFE
NO
k=1000?
YES
CALCULATE
HISTGRAM
STATISTICS
PLOT
HISTGRAM
A
NO
LAST
RUN?
YES
STOP

```
      PROGRAM ROBUST
      DIMENSION NSAMP(5),DATA(5,30),SM(5),SSQ(5),VAR(5),STOT(5)
      DIMENSION NK(5),NNK(5,5)
      COMMON/HISTGRAM/NHIST,XMIN,XMAX,DELX,NBINS
      COMMON DATA,NSAMP,VAR,NS,NK,SM,SSQ
      TYPE REAL MILLERT
      T=TIMEF(T)
      CALL RANFSET(T)
      READ 50,NRUNS
      DO 46 NPASS=1,NRUNS
      READ 50,NS
      READ 60,(NSAMP(I),I=1,NS)
      READ 50,NSCHRUN
      DO 5,NSCH=1,NSCHRUN
      READ 60,(NNK(NSCH,I),I=1,NS)
    5 CONTINUE
      NHIST=1
      XMIN=0.0
      XMAX=100.0
      DELX=1.0
      NBINS=0
      CALL SETLIM(Z,0)
      DO 40 K=1,1000
      DO 10 J=1,NS
      SM(J)=0.0
      SSQ(J)=0.0
```

```
      STOT(J)=0.0
   10 VAR(J)=0.0
      DO 21 J=1,NS
      NNSAMP=NSAMP(J)
      DO 20 I=1,NNSAMP
      CALL NORMAL
      DATA(J,I)=X
      STOT(J)=STOT(J)+X
   20 CONTINUE
      SM(J)=STOT(J)/NNSAMP
   21 CONTINUE
      DO 30 J=1,NS
      NNSAMP=NSAMP(J)
      DO 25 I=1,NNSAMP
      SSQ(J)=(DATA(J,I)-SM(J))**2+SSQ(J)
   25 CONTINUE
   30 VAR(J)=SSQ(J)/(NSAMP(J)-1.0)
      DO 41 NSCH=1,NSCHRUN
      DO 42 M=1,NS
      NK(M)=NNK(NSCH,M)
   42 CONTINUE
      CALL SCHEFE(FSCHEF,DFM,DFD)
      NHIST=1
      CALL HISTSUMS(FSCHEF)
   41 CONTINUE
   40 CONTINUE
      NPLOT=1
```

```
      DO 45 I=1,NPLOT
      CALL CONLEV(I)
      NHIST=I
      CALL HISTPLOT
   45 CONTINUE
   46 CONTINUE
   50 FORMAT(I1)
   60 FORMAT(5I2)
   66 FORMAT(3F10.2)
      END
```

```
      PROGRAM ROBUST
      DIMENSION NSAMP(5),DATA(5,30),SM(5),SSQ(5),VAR(5),STOT(5)
      DIMENSION NK(5),NNK(5,5)
      COMMON/HISTGRAM/NHIST,XMIN,XMAX,DELX,NBINS
      COMMON DATA,NSAMP,VAR,NS,NK,SM,SSQ
      COMMON CL(30)
      TYPE REAL MILLERT
      T=TIMEF(T)
      CALL RANFSET(T)
      READ 50,NRUNS
      DO 46 NPASS=1,NRUNS
      READ 70,(CL(I),I=1,30)
   70 FORMAT(5(F8.3,2X))
      READ 50,NS
      READ 60,(NSAMP(I),I=1,NS)
      READ 50,NSCHRUN
      DO 5,NSCH=1,NSCHRUN
      READ 60,(NNK(NSCH,I),I=1,NS)
    5 CONTINUE
      NHIST=1
      XMIN=0.0
      XMAX=1.0
      DELX=0.02
      NBINS=0
      TT=1.
      CALL SETLIM(Z,TT)
```

```
NHIST=2
XMIN=0.0
XMAX=15.
DELX=0.2
NBINS=0
CALL SETLIM(Z)
NHIST=3
XMIN=0.0
XMAX=8.0
DELX=0.1
NBINS=0
CALL SETLIM(Z)
NHIST=4
XMIN=0.0
XMAX=100.0
DELX=1.0
NBINS=0
CALL SETLIM(Z)
NHIST=5
XMIN=0.0
XMAX=50.0
DELX=0.5
NBINS=0
CALL SETLIM(Z)
NHIST=6
XMIN=-4.0
```

```
      XMAX=4.0
      DELX=0.1
      NBINS=0
      CALL SETLIM(Z)
      DO 40 K=1,1000
      DO 10 J=1,NS
      SM(J)=0.0
      SSQ(J)=0.0
      STOT(J)=0.0
   10 VAR(J)=0.0
      DO 21 J=1,NS
      NNSAMP=NSAMP(J)
      DO 20 I=1,NNSAMP
      IF(J.EQ.2) GO TO 11
      CALL LAPLACE(X)
      GO TO 12
   11 CALL LAPLAC2(X)
   12 CONTINUE
      DATA(J,I)=X
      STOT(J)=STOT(J)+X
   20 CONTINUE
      SM(J)=STOT(J)/NNSAMP
   21 CONTINUE
      DO 30 J=1,NS
      NNSAMP=NSAMP(J)
      DO 25 I=1,NNSAMP
```

```
      SSQ(J)=(DATA(J,I)-SM(J))**2+SSQ(J)
   25 CONTINUE
   30 VAR(J)=SSQ(J)/(NSAMP(J)-1.0)
      CALL COCHRAN(VAR,NS,VARMAX,C)
      NHIST=1
      CALL HISTSUMS(C)
      CALL HARTLEY(VAR,NS,VARMAX,FMAX)
      NHIST=2
      CALL HISTSUMS(FMAX)
      CALL MILLER(TMILLER,MILLERT)
      NHIST=3
      CALL HISTSUMS(TMILLER)
      NHIST=6
      CALL HISTSUMS(MILLERT)
      DO 41 NSCH=1,NSCHRUN
      DO 42 M=1,NS
      NK(M)=NNK(NSCH,M)
   42 CONTINUE
      CALL SCHEFE(FSCHEF,DFM,DFD)
      NHIST=NSCH+3
      CALL HISTSUMS(FSCHEF)
   41 CONTINUE
   40 CONTINUE
      NPLOT=NSCHRUN+3
      DO 45 I=1,NPLOT
      CALL CONLEV(I)
```

```
      CALL POWER(I)
      NHIST=1
      CALL HISTPLOT
   45 CONTINUE
      NHIST=6
      CALL HISTPLOT
   46 CONTINUE
   50 FORMAT(I1)
   60 FORMAT(5I2)
   66 FORMAT(3F10.2)
      END
```

```
      SUBROUTINE CONLEV(IH)
      COMMON/999/IFREQ(102,10),ICOUNT(102),XL(10),XH(10),DX(10),NB(10
     X   IRJCT(10),IUSE(10),IP(100),IX(100) ,FERR(102,10)
      DIMENSION FREQ(102,10)
      EQUIVALENCE(IFREQ,FREQ)
      DIMENSION N(100),VAL(3),L(3)
      B=DX(IH)
      X=XH(IH)
      NBB=NB(IH)
      NR=FREQ(NBB,IH)
      IND=NB(IH)-1
      DO 30 J=1,IND
      I=NB(IH)-J
      N(J)=FREQ(I,IH)
   30 CONTINUE
      NL=3
      L(1)=10
      L(2)=50
      L(3)=100
      LOC=1
  120 LIM=L(LOC)
      LAB=LOC
      CALL POINT(NR,N,LIM,VAL,LAB,X,B)
      LAB=LOC+1
      AL=L(LOC)
      LIM=AL+SQRTF(AL)
```

```
      CALL POINT(NR,N,LIM,VAL,LAB,X,B)
      LAB=LOC+2
      AL=L(LOC)
      LIM=AL-SQRTF(AL)
      CALL POINT(NR,N,LIM,VAL,LAB,X,B)
      CALL PRINT(LOC,VAL)
      LOC=LOC+1
      IF(LOC.GT.NL) GO TO 250
      GO TO 120
  250 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE POINT(NR,N,LIM,VAL,LAB,X,B)
      DIMENSION N(100),VAL(3)
      NSUM=NR
      DO 370 J=1,100
      NSUM=NSUM+N(J)
      INDEX=J
      IF(NSUM.GT.LIM) GO TO 385
  370 CONTINUE
      PRINT 400
  385 ALIM=LIM
      ASUM=NSUM
      AN=N(INDEX)
      ADEX=INDEX-1
      VAL(LAB)=X-B*ADEX-((ALIM+AN-ASUM)/AN)*B
  400 FORMAT(12H POINT ERROR)
      RETURN
      END
```

```
      SUBROUTINE PRINT(LOC,VAL)
      DIMENSION VAL(3)
      PRINT 590,LOC
      PRINT 600,VAL(LOC),VAL(LOC+1),VAL(LOC+2)
  590 FORMAT(5H THE ,I1,20H CONFIDENCE LEVEL IS)
  600 FORMAT(F8.3,2H +,F8.3,2H -,F8.3)
      RETURN
      END
```

```
      SUBROUTINE POWER(IH)
      COMMON/999/IFREQ(102,10),ICOUNT(102),XL(10),XH(10),DX(10),NB(10)
     X   IRJCT(10),IUSE(10),IP(100),IX(100) ,FERR(102,10)
      COMMON DATA,NSAMP,VAR,NS,NK,SM,SSQ
      COMMON CL(30)
      DIMENSION NSAMP(5),DATA(5,30),SM(5),SSQ(5),VAR(5),STOT(5)
      DIMENSION NK(5),NNK(5,5)
      DIMENSION FREQ(102,10)
      EQUIVALENCE(IFREQ,FREQ)
      NBB=NB(IH)
      IND=NBB-1
      DO 40 K=1,6
      BL=XH(IH)
      IC=IH+5*(K-1)
      CLEV=CL(IC)
      NSUM=FREQ(NBB,IH)
      DO 10 I=1,IND
      J=NBB-I
      NSUM=NSUM+FREQ(J,IH)
      BL=BL-DX(IH)
      IF(BL.LT.CLEV) GO TO 20
   10 CONTINUE
   20 BETA=NSUM/1000.
  120 PRINT 30,CLEV,BETA
   30 FORMAT(15H  POWER AT CL =,F8.3,4H  IS,F5.3)
   40 CONTINUE
```

```
RETURN
END
```

I Program Name.

PROGRAM HISTGRAM (x, WT)

II Purpose.

The program accepts data for a maximum of 10 histograms and plots them on line on the printer.

III How to use.

1. Required non-executable statements COMMON / HISTGRAM / NHIST, XMIN, XMAX, DELX, NBINS

COMMON / TITLE / I TITLE

TYPE MANY 7(7) I TITLE

2. Dictionary of common variables.

NHIST - the number of the histogram $1 \leq NHIST \leq 10$

XMIN - minimum bin limit

XMAX - maximum bin limit

DELX - bin size

NBINS - number of bins

ITITLE - histogram title, use 56 H ....

3. Determination of bin size and number of bins. The users have two options.

a) Specify the bin size. Set NBINS = 0. The program will calculate the number of bins. If the number of bins is greater than 102, the bin size is adjusted if possible until the number of bins is less than or equal to 102.

$$NBINS = \frac{XMAX - XMIN}{DELX} + 2$$

b) Specify the number of bins and set DELX = 0.

The program will calculate DELX by:

$$DELX = \frac{XMAX - XMIN}{NBINS - 2}$$

4. Calling Sequence

To set the limits for each histogram, give the appropriate /HISTGRAM/ common variables, then call SETLIM (X, WT). Repeat for each histogram. To insert an element x into a histogram, specify the histogram number (NHIST = n), then call HISTSOMS (X, WT): default for the weighting factor gives WT = 1. To plot the histogram number (NHIST = n), give the title, if any (ITITLE = 56 H ...), then call HISTPLOT (X, WT). Repeat for each histogram.

IV Deck Set Up

```
TIME
JOB
FTN,L,X
CALLING PROGRAM
SUBROUTINE HISTGRAM
   SCOPE
LOAD
RUN
```

V Space Required

Program length	1332 (octal)
COMMON /999/	2534 (octal)

```
      SUBROUTINE HISTGRAM(X,WT)
      COMMON/TITLE/ITITLE(7)
      COMMON/HISTGRAM/N     ,XMIN,XMAX,DELX,NBINS
      EQUIVALENCE (I,IBINS),( ILARG,III),(XX,II),(IUSE,IRJCT)
      COMMON/999/IFREQ(102,10),ICOUNT(102),XL(10),XH(10),DX(10),NB(10
     X    IRJCT(10),IUSE(10),IP(100),IX(100) ,FERR(102,10)
      DIMENSION FREQ(102,10)  $ EQUIVALENCE(IFREQ,FREQ)
      DATA(ITITLE=7(8H         ))
      DATA(IA=0)
      ENTRY SETLIM        $  WT=1.0
       IF(IA.NE.0) GO TO 3
      DO 1 II=1,100
    1  IX(II)=(8R******** )
       DO 2 II=1,1270
    2 IFREQ(II) = 0              $ IA=10
    3 IF( N.LE.10 .AND. N.GT.0 ) GO TO 10
    5 WRITE(61,1000)N           $ RETURN
 1000 FORMAT(///10X,* HISTOGRAM NUMBER * I8,* NOT IN RANGE 1-10 OR I
     1S PREVIOUSLY BEEN USED * /// )
   10 IF ( IUSE(N).EQ. 4RUSED )  GO TO 5
      IUSE(N)=4RUSED     $XL(N)=XMIN     $XH(N)=XMAX  $DX(N)=DELX
      NB(N) = NBINS
      IF ( NB(N)) 50,20,40
   20 IBIN = 0
      IF(XMIN.LT.XMAX) GO TO 21
      WRITE(61,19)XMIN,XMAX   $  GO TO 30
```

```
 19 FORMAT( *  XMIN(*E12.3*) * XMAX(*E12.3*)*)
 21 NB(N)=(XH(N) - XL(N))/DX(N)  + 2.0
    IF( NB(N).LE.102)  RETURN
    IF( IBIN - 10 ) 25,30,30
 25 IBIN =IBIN +1  $  DX(N)= DX(N)*2.0   $ GO TO 21
 30 IRJCT(N) = 6RREJECT
    WRITE(61,1005) N,DELX  $ RETURN
1005 FORMAT(///10X, *HISTOGRAM * I8, * REJECTED ,DELX= * E12.4,* TOO
   1ALL * ///)
 40 IF(NBINS.GT.100) GO TO 50
    DX(N) =   (XH(N)-XL(N))/(NB(N)-2)     $ RETURN
 50 IRJCT(N) = 6RREJECT
    WRITE(61,1010) N,NBINS   $ RETURN
1010 FORMAT (///10X,*HISTOGRAM * I8, * REJECTED, NBINS = * I8,///)
    ENTRY HISTSUMS
    IF ( N.LE.10 .AND. N.GT.0 )  GO TO 100
    WRITE(61,1015) N
1015 FORMAT (///,10X,* HISTSUMS CALLED WITH NHIST = * I8,*OUT OF RAN
   1*,///)
100 IF ( IRJCT(N) .EQ. 6RREJECT )  RETURN
    I =   ((X - XL(N)) / DX(N)) + 2.0
      IF( I-1) 105,105,115
105 I  =  1      $   GO TO 125
115 IF (  I - NB(N) ) 125,125,120
120  I=NB(N)
125  FREQ (I,N) =   FREQ(I,N) + WT
```

```
      FERR(1,N)=FERR'I,N)+WT*WT$ WT=1.0  $  RETURN
      ENTRY HISTPLOT
      IF ( N.LE.10 .AND. N.GT.0 )  GO TO 200
       WRITE(61,1020) N    $ RETURN
 1020 FORMAT(///10X, * HISTPLOT CALLED  WITH NHIST = *,I8, * OUT OF R
     1E * ///)
  200 IF( IRJCT(N) .EQ. 6RREJECT )  GO TO 280
       ICOUNT(1) = ILARG =  FREQ(1,N)
      K = NB(N)
      DO  210   I = 2, K
       ICOUNT(I)  =  FREQ(I,N)
      IF ( ILARG .GE.  FREQ(I,N) ) GO TO 210
      ILARG =  FREQ( I,N)
  210 CONTINUE
      IF ( ILARG .LE. 100 )  GO TO 225
      II =  ILARG /100  + 1
      III=  1
      DO  215  I = 1,K
  215 IFREQ(I,N) =   FREQ(I,N) / II
      GO  TO  240
  225 IF ( ILARG .GT.50 )   GO TO 238
      II =  100 / ILARG
      DO  230  I = 1,K
  230 IFREQ(I,N) =   FREQ(I,N) * II
      III  = II
      II   =  1
```

```
      GO TO  240
 238  II = 1  $  III = I
      DO 239 I=1,K
 239 IFREQ(I,N)=FREQ(I,N)
 240 WRITE(61,1025)N,XL(N),DX(N),NB(N),XH(N)
1025 FORMAT(1H1,40X,* HISTOGRAM PLOT---NHIST = *I3/13X,*XMIN= *E12.4
    Z5X,*DELTX= *,E12.4, * BINS= *,I4,5X,*XMAX= *E12.4)
      WRITE(61,1030)III,II
1030 FORMAT(/ 20X,  I8, * ASTERISKS  =  * ,I8 ,* COUNTS--- MULTIPLY
    1DIVIDE ACCORDINGLY *)
      WRITE(61,1031)ITITLE
1031  FORMAT(/,20X,7A8)
      WRITE(61,1035)
1035 FORMAT(/ 23X,1H0,4X,1H5,4X,2(1H1, 4X), 2(1H2, 4X ), 2(1H3, 4X)
    12(1H4,4X) , 2( 1H5, 4X) ,  2(1H6, 4X), 2(1H7, 4X ), 2(1H8, 4X)
    12(1H9,4X) ,1H1,/ 33X, 9(1H0,4X,1H5,4X),1H0,/10X, 113X,1H0,7X)
      WRITE(61,1040)
1040 FORMAT(3X,10HERROR      ,10H   COUNT    , 100(1H-))
      XX =  -10.E-299
      WRITE(61,1045) XX
1045  FORMAT(11X,E11.3,X,1H- )
      XX =  XL(N) - DX(N)
        ISUM=0
      DO  270 I = 1,K,I
      DO  245 III =1,100
 245 IP(III) = ( 8R          )
```

```
      ERR = SQRTF(FERR(I,N)   )
      III =  IFREQ(I,N)
        IFREQ(I,N) = 0
      FERR(I,N)=0
          DO 248 J=1,III
 248    IP(J)  = IX(J)
       ISUM = ISUM + ICOUNT( I)
      WRITE(61,1056) ERR,ICOUNT(I),(IP(III),III=1,100)
       ICOUNT(I) = (8R          )
            IF(K.GE.40) GO TO 253
      WRITE(61,1059)          (IP(III),III=1,100)
 1059 FORMAT(23X,1H-,100R1)
 1056 FORMAT(1X,F10.3,    I11,2H -, 100R1 )
  253   XX=XX+DX(N)
          IF( I.EQ.K) XX     =1.E+300
      WRITE(61,1057)XX,(IP(III),III=1,100)
 1057  FORMAT(  11X,         E11.3,2H -,100R1)
  270 CONTINUE
       WRITE(61,1060) ISUM
 1060 FORMAT(/3X,*TOTAL COUNT=*I10)
      DO 1058 I=1,7
 1058 ITITLE(I)=8H
  280           IUSE(N)=0$RETURN$END
```

```
      SUBROUTINE COCHRAN(VAR,NS,VARMAX,C)
      DIMENSION VAR(5)
      VARMAX=VAR(1)
      SUMVAR=0.0
      NNS=NS-1
      DO 120 I=1,NNS
      IF(VARMAX-VAR(I+1)) 110,120,120
  110 VARMAX=VAR(I+1)
  120 CONTINUE
      DO 130 I=1,NS
  130 SUMVAR=SUMVAR+VAR(I)
      C=VARMAX/SUMVAR
      RETURN
      END
```

```
      SUBROUTINE HARTLEY(VAR,NS,VARMAX,FMAX)
      DIMENSION VAR(5)
      VARMIN=VAR(1)
      NNS=NS-1
      DO 220 I=1,NNS
      IF(VARMIN-VAR(I+1)) 220,210,210
  210 VARMIN=VAR(I+1)
  220 CONTINUE
      FMAX=VARMAX/VARMIN
      RETURN
      END
```

```
      SUBROUTINE MILLER(TMILLER,MILLERT)
      DIMENSION NK(5),SM(5),SSQ(5)
      DIMENSION NSAMP(5),DATA(5,30),VAR(5),DATAJK(30),JLGVAR(5,30)
      DIMENSION SJLGVAR(5),TXI(30),TYI(30)
      DIMENSION SJVAR(5,30)
      COMMON DATA,NSAMP,VAR,NS,NK,SM,SSQ
      COMMON CL(30)
      TYPE REAL MILLERT
      TYPE REAL JLGVAR
      TYPE REAL N,M
      TYPE REAL NN
      DO 30 I=1,NS
      SJLGVAR(I)=0.0
      NNSAMP=NSAMP(I)
      DO 30 K=1,NNSAMP
      JLGVAR(I,K)=0.0
      SJVAR(I,K)=0.0
      DO 2 JJ=1,NNSAMP
    2 DATAJK(JJ)=0.0
      SSQTX=0.0
      SSQTY=0.0
      NA=0
      SUMDJK=0.0
      AVJKD=0.0
      L=0
    5 DO 20 J=1,NNSAMP
```

```
      IF(J-K)15,10,15
   10 NA=-1
      GO TO 20
   15 L=J+NA
      DATAJK(L)=DATA(I,J)
      SUMDJK=SUMDJK+DATAJK(L)
   20 CONTINUE
      NNI=NNSAMP-1
      NN=NNI
      AVJKD=SUMDJK/NN
      DO 25 J=1,NNI
      SJVAR(I,K)=(((DATAJK(J)-AVJKD)**2)/NN)+SJVAR(I,K)
      JLGVAR(I,K)=LOGF(SJVAR(I,K))
   25 CONTINUE
      SJLGVAR(I)=SJLGVAR(I)+JLGVAR(I,K)
   30 CONTINUE
      N=NSAMP(1)
      M=NSAMP(2)
      NNN=NSAMP(1)
      MMM=NSAMP(2)
      TXPOP=0.0
      TYPOP=0.0
      TX=LOGF(VAR(1))
      TY=LOGF(VAR(2))
      TDX=N*TX-((N-1.0)/N)*SJLGVAR(1)
      TDY=M*TY-((M-1.0)/M)*SJLGVAR(2)
```

```
      DO 40 J=1,NNN
      TXI(J)=N*TX-(N-1.0)*JLGVAR(1,J)
   40 CONTINUE
      DO 50 J=1,MMM
      TYI(J)=M*TY-(M-1.0)*JLGVAR(2,J)
   50 CONTINUE
      DO 60 J=1,NNN
      SSQTX=SSQTX+(TXI(J)-TDX)**2
   60 CONTINUE
      DO 70 J=1,MMM
      SSQTY=SSQTY+(TYI(J)-TDY)**2
   70 CONTINUE
      A=(TDX-TXPOP)-(TDY-TYPOP)
      B=(1.0/(N*(N-1.0)))*SSQTX+(1.0/(M*(M-1.0)))*SSQTY
      MILLERT=A/SQRT(B)
      TMILLER=ABSF(MILLERT)
      RETURN
      END
```

```
      SUBROUTINE SCHEFE(FSCHEF,DFN,VE)
      COMMON DATA,NSAMP,VAR,NS,NK,SM,DUM
      DIMENSION VAR(5),SM(5),DUM(5)
      DIMENSION DATA(5,30),NSAMP(5),SDATA(5,5,30),NK(5),KCS(5,5)
      DIMENSION SUMX(5,5),SSQ(5,5),Y(5,5),ATAV(5),CCSV(5)
      DO 30 I=1,NS
      L=1
      NNSAMP=NSAMP(I)
      ANNSAMP=NNSAMP
      ANK=NK(I)
      NR=(ANNSAMP/ANK)+0.5
      DO 30 J=1,NNSAMP
      NTEST=(J)-NR*L
      IF(NTEST) 20,20,10
   10 L=L+1
   20 M=(J)-NR*(L-1)
      SDATA(I,L,M)=DATA(I,J)
   30 KCS(I,L)=M
      DO 40 I=1,NS
      NNK=NK(I)
      DO 40 J=1,NNK
      SUMX(I,J)=0.0
      SSQ(I,J)=0.0
   40 CONTINUE
      DO 70 I=1,NS
      NNK=NK(I)
```

```
      DO 70 J=1,NNK
      KKCS=KCS(I,J)
      DO 50 K=1,KKCS
      SUMX(I,J)=SDATA(I,J,K)+SUMX(I,J)
   50 CONTINUE
      DO 60 K=1,KKCS
      Q=((SDATA(I,J,K)-SUMX(I,J)/KCS(I,J))**2)/(KCS(I,J)-1.0)
      SSQ(I,J)=SSQ(I,J)+Q
   60 CONTINUE
      Y(I,J)=LOGF(SSQ(I,J))
      ATAV(I)=0.0
      CCSV(I)=0.0
   70 CONTINUE
      VE=0.0
      SUMVY2=0.0
      SUMVN2=0.0
      CCSS=0.0
      ATAS=0.0
      AMAX=NS
      DO 90 I=1,NS
      AJMAX=NK(I)
      VE=VE+AJMAX-1.0
      NNK=NK(I)
      DO 80 J=1,NNK
      CCSV(I)=CCSV(I)+KCS(I,J)-1.0
      SUMVY2=SUMVY2+(KCS(I,J)-1.0)*Y(I,J)**2
```

```
80 CONTINUE
   DO 85 J=1,NNK
   ATAV(I)=ATAV(I)+(KCS(I,J)-1.0)*Y(I,J)/CCSV(I)
85 CONTINUE
   CCSS=CCSS+CCSV(I)
90 CONTINUE
   DO 95 I=1,NS
   ATAS=ATAS+ATAV(I)*CCSV(I)/CCSS
   SUMVN2=SUMVN2+CCSV(I)*(ATAV(I)**2)
95 CONTINUE
   FSCHEF=(VE*(SUMVN2-CCSS*ATAS**2))/((AMAX-1.0)*(SUMVY2-SUMVN2))
   DFN=AMAX-1.0
   RETURN
   END
```

```
      SUBROUTINE BOXAND(D)
      DIMENSION DATA(5,30),NSAMP(5),SM(5),SSQ(5),VAR(5),DEV4(5)
      DIMENSION NK(5)
      COMMON DATA,NSAMP,VAR,NS,NK,SM,SSQ
      TYPE REAL N
      DO 10 I=1,2
      DEV4(I)=0.0
      NNSAMP=NSAMP(I)
      DO 10 J=1,NNSAMP
      DEV4(I)=(DATA(I,J)-SM(I))**4+DEV4(1)
   10 CONTINUE
      A=DEV4(1)+DEV4(2)
      B=(SSQ(1)+SSQ(2))**2
      N=NSAMP(1)+NSAMP(2)
      B2=(N+2)*(A/B)
      D=1.0/(1.0+.5*((N+2)/(N-1.0-(B2-3.0)))*(B2-3.0))
      R=VAR(1)-VAR(2)
      IF(R) 20,20,30
   20 F=VAR(2)/VAR(1)
      DFN=D*(NSAMP(2)-1)
      DFD=D*(NSAMP(1)-1)
      GO TO 40
   30 F=VAR(1)/VAR(2)
      DFN=D*(NSAMP(1)-1)
      DFD=D*(NSAMP(2)-1)
   40 CONTINUE
```

```
      IDFN=DFN+.5
      IDFD=DFD+.5
      PUNCH 60,F,IDFN,IDFD
   60 FORMAT(3X,2HF=,F6.3,6H  DFN=,I3,6H  DFD=,I3)
      RETURN
      END
```



```
   60 FORMAT(3X,2HF=,F6.3,6H  DFN=,I3,6H  DFD=,I3)
      RETURN
      END
```

```
      SUBROUTINE NORMAL(X)
      X=0.0
      DO 90 I=1,12
   90 X=RANF(-1)+X
      X=X-6.0
      RETURN
      END
```

```
      SUBROUTINE NORMAL5(X)
      X=0.0
      DO 90 I=1,60
   90 X=RANF(-1)+X
      X=X-30.0
      RETURN
      END
```

```
SUBROUTINE RECTANG(X)
X=0.0
U=RANF(-1)
X=-1.732+3.464*U
RETURN
END
```

```
SUBROUTINE RECT2(X)
X=0.0
U=RANF(-1)
X=-3.873+7.746*U
RETURN
END
```

```
      SUBROUTINE LAPLACE(X)
      ALPHA=-4.0
      BETA=4.0
      B=1.0/1.4142
      F=1.0/(2*B)
      X=0.0
    1 U1=RANF(-1)
      U2=RANF(-1)
      X=ALPHA+(BETA-ALPHA)*U2
      IF(X) 2,3,3
    2 Y=-X
      GO TO 4
    3 Y=X
    4 FY=(1.0/(2.0*B))*(EXPF((-Y)/B))
      IF((FY/F)-U1) 1,1,5
    5 CONTINUE
      RETURN
      END
```

```
  SUBROUTINE LAPLAC2(X)
  ALPHA=-4.0
  BETA=4.0
  B=1.581139
  F=1.0/(2*B)
  X=0.0
1 U1=RANF(-1)
  U2=RANF(-1)
  X=ALPHA+(BETA-ALPHA)*U2
  IF(X) 2,3,3
2 Y=-X
  GO TO 4
3 Y=X
4 FY=(1.0/(2.0*B))*(EXPF((-Y)/B))
  IF((FY/F)-U1) 1,1,5
5 CONTINUE
  RETURN
  END
```

```
      SUBROUTINE DEXP(X)
      R1=-4.0
      R2=4.0
      B=2.449490/3.141593
      A=-0.5772156*B
      F=(1.0/B)*0.36787944
    1 U1=RANF(-1)
      U2=RANF(-1)
      X=R1+(R2-R1)*U2
      Y=(X-A)/B
      DE=(1/B)*EXPF(-Y-EXPF(-Y))
      IF((DE/F)-U1)1,1,5
    5 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE DEXP5(X)
      R1=-4.0
      R2=4.0
      B=5.477266/3.141593
      A=-0.5772156*B
      F=(1.0/B)*0.36787944
    1 U1=RANF(-1)
      U2=RANF(-1)
      X=R1+(R2/R1)*U2
      Y=(X-A)/B
      DE=(1/B)*EXPF(-Y-EXPF(-Y))
      IF((DE/F)-U1)1,1,5
    5 CONTINUE
      RETURN
      END
```