NATIONAL BUREAU OF STANDARDS REPORT

9639

SOME PROPERTIES OF FOUR

COCHRAN-TYPE TESTS FOR HOMOGENEITY OF VARIANCE

NBS

U.S. DEPARTMENT OF COMMERCE

NATIONAL BUREAU OF STANDARDS

THE NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards[1] provides measurement and technical information services essential to the efficiency and effectiveness of the work of the Nation's scientists and engineers. The Bureau serves also as a focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. To accomplish this mission, the Bureau is organized into three institutes covering broad program areas of research and services:

THE INSTITUTE FOR BASIC STANDARDS . . . provides the central basis within the United States for a complete and consistent system of physical measurements, coordinates that system with the measurement systems of other nations, and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. This Institute comprises a series of divisions, each serving a classical subject matter area:

—Applied Mathematics—Electricity—Metrology—Mechanics—Heat—Atomic Physics—Physical Chemistry—Radiation Physics—Laboratory Astrophysics[2]—Radio Standards Laboratory,[2] which includes Radio Standards Physics and Radio Standards Engineering—Office of Standard Reference Data.

THE INSTITUTE FOR MATERIALS RESEARCH . . . conducts materials research and provides associated materials services including mainly reference materials and data on the properties of materials. Beyond its direct interest to the Nation's scientists and engineers, this Institute yields services which are essential to the advancement of technology in industry and commerce. This Institute is organized primarily by technical fields:

—Analytical Chemistry—Metallurgy—Reactor Radiations—Polymers—Inorganic Materials—Cryogenics[2]—Office of Standard Reference Materials.

THE INSTITUTE FOR APPLIED TECHNOLOGY . . . provides technical services to promote the use of available technology and to facilitate technological innovation in industry and government. The principal elements of this Institute are:

—Building Research—Electronic Instrumentation—Technical Analysis—Center for Computer Sciences and Technology—Textile and Apparel Technology Center—Office of Weights and Measures —Office of Engineering Standards Services—Office of Invention and Innovation—Office of Vehicle Systems Research—Clearinghouse for Federal Scientific and Technical Information[3]—Materials Evaluation Laboratory—NBS/GSA Testing Laboratory.

---

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D. C., 20234.

[2] Located at Boulder, Colorado, 80302.

[3] Located at 5285 Port Royal Road, Springfield, Virginia 22151.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT | **NBS REPORT**

20503-12-2050131 | November 7, 1967 | 9639

SOME PROPERTIES OF FOUR

COCHRAN-TYPE TESTS FOR HOMOGENEITY OF VARIANCE

by

Brian L. Joiner

NBS

U.S. DEPARTMENT OF COMMERCE

NATIONAL BUREAU OF STANDARDS

TABLE OF CONTENTS

Contents | Page

TABLE OF CONTENTS (Continued)

Contents | Page

1. INTRODUCTION

Cochran (1941) suggested using the test statistic $S^2_{max} / \sum_{j=1}^{k} S^2_j$ "in testing one of a group of estimates of variance which appears to be anomalously large." In this study we investigate some properties of four such tests all of the general form

$$V_{max} / \sum_{j=1}^{k} V_j$$

where the V_j are independent estimates of dispersion. In these four tests the V_j are respectively: (1) sample variances (Cochran, 1941), (2) sample ranges (Bliss, Cochran and Tukey, 1956), (3) sample standard deviations (suggested by Churchill Eisenhart), and (4) sample mean deviations. The emphasis throughout is on small sample properties.

This study was motivated by a desire to find a simple test for homogeneity of variance that would not be as sensitive to non-normality as those in the literature, and to provide some quantative comparisons among some of those already available. A brief discussion of the conclusions of the study is given in Chapter 2, and two suggested approaches toward more robust tests for homogeneity of variance are given in Chapter 3. These two approaches are suggested by the results we have obtained in this work and may be considered topics for further research.

Chapter 4 consists of a collection of information on the small sample properties of the standard deviation, mean deviation and range in samples from non-normal distributions. Information available in the literature on the expectation and coefficient of variation of these three estimators is summarized in Section 4.3. New information on the three estimators is given in Section 4.5 for samples of size 3, 5, and 10 from a number of discrete distributions each having at most eleven categories. In addition new information on the range is given in Section 4.4 for samples from a family of random variables suggested by Tukey. All of this information is compared with an approximation to the lower moments of the standard deviation based on a modified degrees of freedom approach suggested by Le Roux (1931). His method seems to provide a very good approximation to the first two moments of the standard deviation, and an adequate approximation for very small sample sizes to the first two moments of the mean deviation. The results on the range are compared with those of Cox (1954) and others in Section 4.6. In Section 4.7 we suggest a modified degrees of freedom approximation for R and give values of the modified sample size for Tukey random variables and Cox's "average."

Some general properties of Cochran-type tests are considered in Chapter 5, and methods for computing approximate percentage points of such tests are described. Chapter 5 also includes a brief outline of the historical development of this class of tests.

In Chapter 6 approximate percentage points are computed for the two new tests based on the standard deviation and mean deviation respectively. For the test based on the standard deviation, called Eisenhart's test, approximate percentage points are given for significance levels of .01, .05 and .10. For the tests based on the mean deviation critical values are given for a significance level of .05.

Chapter 7 consists of two sections, the first of which considers some properties of the four tests under normality and the second of which considers some approximate properties of the four tests under non-normality. In the first section a single alternative to the null hypothesis of equal variance is considered. In this alternative an unknown one of the underlying variances is assumed to have been inflated in the ratio $\theta^2 > 1$. Two functions closely related to the power function are introduced and some numerical values computed. Let P(RC) denote the probability that the null hypothesis is rejectable for the correct reason, that is, the probability that $V_i / \Sigma V_j$ is larger than the critical value when the i'th population is indeed the one with the inflated variance. Then it is shown that P(RC) constitutes the bulk of the power function for at least moderately distant alternatives. Some numerical values are computed for P(RC) and for a modified version of the "median significance level."

The "median significance level" (MSL) is introduced in Appendix B as an alternative to the power function as a means of assessing the properties of statistical tests. The MSL may be simply described as the median of the distribution of the observed significance level for a given alternative hypothesis. It is a single function that represents some of the properties of a test that are given in full by the family of power functions. Some elementary examples are given in which the median significance level is compared with several similar criteria.

In Appendix A several approximations to the distribution of the sum of independent chi variates are compared. These approximations are used in the computation of percentage points of Eisenhart's test in Section 6.1.

2. SUMMARY AND CONCLUSIONS ON COCHRAN-TYPE TESTS

In this chapter we summarize our principal conclusions concerning tests for homogeneity of variance. Summaries of detailed results are given in several other places; in particular, results on the expectation and coefficient of variation of S, M, and R are summarized in Sections 4.5 and 4.6, and results on properties of four Cochran-type tests under normality and under non-normality are summarized in Sections 7.1 and 7.2, respectively.

Our approach has differed from previous work in this area in that we have considered only small sample properties. In addition we have focused attention on a particular class of tests and compared the properties of several tests in this class. The choice of the particular class of tests (Cochran-type tests) was based primarily on its intuitive appeal since tests in this class would appear to be sensitive to alternative hypotheses that are frequently of interest. In addition, this class of tests has the desirable properties of being relatively easy to compute and relatively easy to interpret.

It may be noted that we have not considered the possibility of structure among the variances. If there were evidence that the variances were related to some variable in a systematic manner, a completely different approach would be pertinent. Bechofer (1960) has, for example, considered the case where the variances are related according to a completely multiplicative model.

We have compared some power-type properties of four tests in this class in Section 7.1 and have found that for the small values of n and k considered here there is virtually no loss of power from using Eisenhart's test rather than Cochran's test. In the class of alternative considered here, an unknown one of the variances has been inflated in the ratio $\theta^2 > 1$. The mean deviation test seems to be slightly less powerful than Eisenhart's test and the BCT test slightly less powerful than the mean deviation test. Bartlett's test is appreciably less powerful than Cochran's test for the type of alternative considered here. This alternative is particularly tailored to be favorable to Cochran-type tests, however, and Bartlett's test would probably prone to be better against other alternatives.

However, our primary emphasis has not been on power-type properties, but on the robustness properties of the various tests when the parent distributions are not normal. Box (1953a) in his fundamental paper on the general lack of robustness of tests for homogeneity of variance has shown that "a test of variances should be 'studentized' for the fourth moment just as a test of means is studentized for the second moment," and has concluded that all tests "which do not utilize evidence on variance variability within the samples are equally sensitive to non-normality." The four tests considered in this work are <u>not</u> studentized and we have shown in Section 7.2 that they are all quite radically effected by non-normality. However in Section 7.2 we have also given some evidence that there may be appreciable differences in the extent to which the various non-studentized tests are affected by different kinds of non-normality.

The evidence given there is not clear-cut since the results based on modified degrees of freedom approximations differ appreciably from the results obtained in the limited sampling study. It appears that these differences may be due primarily to differences between parent distributions all having the same value of β_2, since Le Roux (1931) has shown that there do exist distributions for which the sample variance does behave much as if it were based on a sample of a modified size from the normal distribution.

It is not easy to draw conclusions in such a situation but if a choice among the tests had to be made based on the robustness results in Section 7.2, it seems doubtful that Bartlett's test could be recommended. The tests based on the mean deviation might be the safest bet. For samples from distributions like the TRV-small distributions, the tests based on the range may be slightly more robust than Cochran's and Eisenhart's tests. Cox's (1954, p.479) comment that "comparative tests based on ranges of small samples are appreciably less affected by non-normality than the corresponding tests based on mean squares," seems to be a slight overstatement but does seem to point in the right direction.

3. TWO SUGGESTED APPROACHES TOWARD MORE ROBUST TESTS

Here we briefly describe two "new" ways of testing for homogeneity of variance that are suggested by the results we have obtained, and that seem intuitively appealing to us. The first of these methods enables one to use almost any conventional test for homogeneity of variance and, to a first order of approximation, to assess the result using critical values appropriate for several different values of the standardized fourth moment β_2 of the parent distribution. Tables are provided. The second "method" is a proposed new test based on the statistic S_k / S_{k-1} where S_k and S_{k-1} denote the largest and second largest observed standard deviations. The properties of the test have not been studied, but heuristic reasons are given why this test is conjectured to be of greater robustness than most of those presently available.

First we mention two existing tests that are apparently quite robust with respect to departures from normality.

Box (1953a), Scheffe (1959, p.83) and Odeh and Olds (1959) have considered a test based on the analysis of variance of logarithms of sample variances. However, this test can only be employed by creating artificial subsamples if one does not wish to assume that any subsets of the σ_i^2 $(i = 1, \ldots, k)$ are equal. In such a case the result obtained in a particular application would have the intuitively unattractive feature of being dependent upon how the artificial subsamples were created.

Box and Andersen (1955) have suggested another procedure in which Bartlett's (1937) test is "studentized" by the observed fourth moment. This procedure does not seem to have been widely accepted, however, perhaps because of the rather heavy computations involved or possibly because of the notorious inaccuracy of samll sample estimates of the fourth moment.

A Simple Approximate Procedure for Compensating for Non-Normality

The method described here consists of treating the sample variances (or possibly other measures of dispersion) as though they were based on samples of a modified size from the normal distribution, the modification depending on the value of β_2 for the parent distributions. Using this method the test statistic is first computed and then assessed against critical values for several modified sample sizes.

Much of the background for this suggestion is described in Sections 4.5, 4.7 and 7.2. The basic idea stems from an observation of Le Roux (1931) and has been used in connection with tests for homogeneity of variance by Box (1953a) and in Section 7.2 of this work.

The basis for the method may briefly be described as follows:

1. The properties of all known tests for homogeneity of variance based on equal sized samples do not depend on the expectation of the sample variances, merely their relative expectations.
2. The properties of tests for homogeneity of variance that are not "studentized" for the parent fourth moment β_2 do depend critically on the coefficient of variation of the variance which in turn is a simple function of β_2 .
3. The coefficient of variation of the sample variance based on a sample of size n from any distribution with standardized fourth moment β_2 , is the same as that for some sample size n^* from the normal distribution. The value of n^* is a function of n and β_2 and may be fractional.
4. Thus the properties of a test for homogeneity of variance based on samples of (equal) size n from distributions with a given value of β_2 , should be at least roughly the same as those for the same test based on samples of size n^* from the normal distribution.

Table 3.1 gives some values for the modified degrees of freedom $\nu^* = n^* - 1$ for all combinations of $\nu = 1(1)10(2)20(5)50(10)100$ and $\beta_2 = 2(1)9$ where ν is the nominal number of degrees of freedom in each sample. Strictly speaking ν^* is exact only for the case where the sample variance is based on $\Sigma(X_i - \bar{X})^2$ in which case $\nu = n-1$, although it may be approximately correct in more general situations.

When β_2 is known this approximate method may readily be used to assess a result obtained from Cochran's, Eisenhart's or Hartley's test. For Bartlett's test the statistic L should be multiplied by ν^*/ν before being compared with the appropriate critical value. As an example let $n = 10$ and $k = 10$, then $\nu = 9$ and the approximate .05 critical values of Cochran's test would be roughly as follows:

β_2	2	3	4	5	6	9
ν^*	16.4	9	6.2	4.7	3.8	2.4
Critical Value	.203	.244	.279	.311	.340	.416

Hence an observed value of the test statistic of .30 would be significant at the .05 level if β_2 were less than about 4.5.

The basis for the above approximate method of compensating for non-normality is implicit in a statement of Box (1953a, p.330). In discussing possible interpretations of the result of Bartlett's test applied to an example of Bartlett and Kendall's he notes that (to the extent of the approximation) the null hypothesis is

Table 3.1

Values of the "modified degrees of freedom" ν^*, for S^2

based on a sample of size $n = \nu+1$ from a distribution with standardized fourth moment β_2.

ν	$\beta_2 = 2$	3	4	5	6	7	8	9	10
1	1.3	1.0	.8	.7	.6	.5	.4	.4	.4
2	3.0	2.0	1.5	1.2	1.0	.9	.8	.7	.6
3	4.8	3.0	2.2	1.7	1.4	1.2	1.0	.9	.8
4	6.7	4.0	2.9	2.2	1.8	1.5	1.3	1.2	1.1
5	8.6	5.0	3.5	2.7	2.2	1.9	1.6	1.4	1.3
6	10.5	6.0	4.2	3.2	2.6	2.2	1.9	1.7	1.5
7	12.4	7.0	4.9	3.7	3.0	2.5	2.2	1.9	1.7
8	14.4	8.0	5.5	4.2	3.4	2.9	2.5	2.2	1.9
9	16.4	9.0	6.2	4.7	3.8	3.2	2.8	2.4	2.2
10	18.3	10.0	6.9	5.2	4.2	3.5	3.1	2.7	2.4
12	22.3	12.0	8.2	6.2	5.0	4.2	3.6	3.2	2.8
14	26.2	14.0	9.5	7.2	5.8	4.9	4.2	3.7	3.3
16	30.2	16.0	10.9	8.2	6.6	5.6	4.8	4.2	3.7
18	34.2	18.0	12.2	9.2	7.4	6.2	5.3	4.7	4.2
20	38.2	20.0	13.5	10.2	8.2	6.9	5.9	5.2	4.6
25	48.1	25.0	16.9	12.7	10.2	8.6	7.3	6.4	5.7
30	58.1	30.0	20.2	15.2	12.2	10.2	8.8	7.7	6.8
35	68.1	35.0	23.6	17.7	14.2	11.9	10.2	8.9	7.9
40	78.1	40.0	26.9	20.2	16.2	13.6	11.6	10.2	9.1
45	88.1	45.0	30.2	22.7	18.2	15.2	13.1	11.4	10.2
50	98.1	50.0	33.6	25.2	20.2	16.9	14.5	12.7	11.3
60	118.1	60.0	40.2	30.2	24.2	20.2	17.3	15.2	13.5
70	138.1	70.0	46.9	35.2	28.2	23.6	20.2	17.7	15.7
80	158.0	80.0	53.6	40.2	32.2	26.9	23.1	20.2	18.0
90	178.0	90.0	60.2	45.2	36.2	30.2	25.9	22.7	20.2
100	198.0	100.0	66.9	50.2	40.2	33.6	28.8	25.2	22.4

not contradicted at the 5% level for any value of β_2 greater than 3.44. Thus one may approximately assess what effect a given degree of non-normality would have on the significance attached to an observed result. If the significance or non-significance of the result may be ascertained independently over a range of non-normality deemed appropriate, then one may be confident of the result. However when these conditions are not realized one is warned to proceed with caution.

When β_2 is not known, Box and Anderson (1955) have suggested that one may use

$$\hat{\beta}_2 = \sum_{j=1}^{k} k_{4j} \Big/ \Big(\sum_{j=1}^{k} k_{2j} \Big)^2$$

where k_{ij} denotes the i'th k-statistic (see Kendall and Stuart, 1958, p. 277) based on the j'th sample. It would, of course, always be preferable if some prior information on β_2 were available.

The validity of this method of approximation has not been ascertained and evidence has been given in Section 7.2 that this method may overcompensate in some cases and undercompensate in others. However it is believed that this approximate method of compensating for the effect of non-normality is better than none at all. It should serve adequately as a qualitative guide in assessing the significance of an observed result in the sense that if the result is significant under strict normality but not under a modest degree of non-normality such as $\beta_2 = 4$ or 5, then little reliance can ordinarily be placed on the result.

Proposed Test Based on S_k / S_{k-1}

In this section we propose a new test which we conjecture to be more robust than those currently available. Heuristic reasons are given in support of our conjecture.

Let $T^{(\ell)} = S_k / S^{(\ell)}$ where the S's are independent sample standard deviations with subscripts assigned in order of increasing magnitude, and where $S^{(\ell)}$ denotes the ℓ'th mean of the k-1 smallest S's. That is, let $S^{(\ell)}$ be defined by

$$S^{(\ell)} = \left(\frac{1}{k-1} \sum_{j=1}^{k-1} S_j^{\ell} \right)^{1/\ell} .$$

Consider the class of tests defined by $T^{(\ell)}$ such that $-\infty \leq \ell \leq \infty$. For finite ℓ, this class of tests is equivalent to a subclass of Cochran-type tests, as defined in Chapter 5; namely the subclass based on powers of the sample variance. For example, $T^{(2)}$ and $T^{(1)}$ are equivalent to Cochran's and Eisenhart's tests, respectively. The limiting case $\ell = -\infty$ yields Hartley's test, S_k/S_1 .

A similar notation and method of classification of tests for homogeneity of variance has been considered by Laue (1965). He considered the class of tests defined by $M^{(\ell)} / M^{(j)}$ where $M^{(\ell)}$

denotes the ℓ'th mean of all of the S's. These means have been considered in more generality by Norris (1935). For finite ℓ, $T^{(\ell)}$ is equivalent to $M^{(\infty)} / M^{(\ell)}$.

Three limiting cases of the ℓ'th mean are described by

$$S^{(-\infty)} = \lim_{\ell \to -\infty} S^{(\ell)} = S_1 ,$$

$$S^{(0)} = \lim_{\ell \to 0} S^{(\ell)} = \left(\prod_{i=1}^{k-1} S_j \right)^{1/k-1} ,$$

and

$$S^{(\infty)} = \lim_{\ell \to \infty} S^{(\ell)} = S_{k-1} .$$

Hartley's (1950) test is equivalent to $T^{(-\infty)}$, and we conjecture that $T^{(\infty)}$ is probably more robust overall than the other tests based on $T^{(\ell)}$. The heuristic argument behind this conjecture may be stated as follows. The numerator of $T^{(\ell)}$ is obviously quite sensitive to non-normality and the most robust test would therefore be one in which the denominator $S^{(\ell)}$ would be as nearly as possible equally sensitive to non-normality, in a compensatory manner. It would seem that the distribution of S_{k-1} would be affected more nearly in the same way as that of S_k than would that of any combination of the S_j for $j \leq k-1$, for most types of non-normality. Thus it seems that the robustness of $T^{(\ell)}$ might increase with ℓ. Conflicting evidence is given in Section 7.2 with respect to the argument that the robustness of $T^{(\ell)}$ is an increasing function of ℓ. First, some evidence in support of the argument is given in the sense that it appears that $T^{(2)}$ (Cochran's test) is slightly more robust than $T^{(1)}$ (Eisenhart's test), but then evidence against the argument's holding in complete generality is given when the sampling study indicates that $T^{(-\infty)}$ (Hartley's test) is slightly more robust overall than either $T^{(2)}$ or $T^{(1)}$. Percentage points of $T^{(\infty)}$ $T^{(\infty)} = S_k / S_{k-1}$ are not presently available so it is not possible to begin a direct evaluation of its robustness, but the intuitive appeal of the test and the heuristic argument above would appear sufficient to warrent a more detailed examination of the test's properties.

4. ROBUSTNESS PROPERTIES OF S, M, AND R .

Conflicting claims have been made as to the relative merits of the sample standard deviation, mean deviation, and range as measures of dispersion when sampling from non-normal populations. For example, physicists, astronomers and other scientists have long held that the mean deviation M is a better measure of dispersion in practice since it gives less weight to the possibly spurious observations that sometimes occur in the tails of a sample. Cox (1954, 1955) has given evidence that certain properties of the range R, are less sensitive to non-normality than are the corresponding properties of the standard deviation S. The standard deviation has long received the blessings of statisticians, although its complete superiority has been rigorously demonstrated only for the normal distribution. The mean deviation has indeed received few kind words from statisticians since it was thrown into disrepute by Fisher (1920). For exceptions, see Tukey (1960) and Herrey (1965).

It seems appropriate, in light of these conflicting statements and of the importance of the question, to pursue in detail a comparative study of the small sample properties of these most commonly used measures of dispersion.

The present investigation is based on comparisons of functions of the first two moments of the sampling distributions of the estimators. In Section 4.1 the basis for the choice of the particular properties investigated is given along with a discussion of the concept of "robustness of scale parameters." Section 4.2 gives a survey of the properties of S, M, and R for samples from the normal distribution. The material in these first two sections serves as background information for the remainder of the Chapter.

In Sections 4.3, 4.4, and 4.5 some properties of S, M, and R are given for a number of non-normal distributions. Section 4.3 consists of what is believed to be a comprehensive survey of the literature on the expectation and coefficient of variation of S, M, and R from non-normal distributions. In Section 4.4 some properties of R are given for samples from the family of distributions known as "Tukey random variables" and in Section 4.5 the expectation and coefficient of variation of S, M, and R are given for samples from a number of discrete distributions. These discrete distributions are not related to the classical discrete distributions but are such that only a small number of equally-spaced distinct values are possible. The number of possible values ranges from 2 to 11. The results on S and M are summarized at the end of the section.

Section 4.6 consists of a summary of the results on R contained in Sections 4.3, 4.4, and 4.5, and a comparison of these results with those already available in the literature. In Section 4.7 we suggest a modified degrees of freedom approach for R and give values of the modified sample size for Tukey random variables and Cox's "average."

4.1 PROPERTIES CONSIDERED

The different properties of an estimator assume varying importance depending upon the purposes for which the estimator is to be used. Three basic uses for estimators of scale parameters are : (1) as a point estimate of a scale parameter, (2) in tests of hypotheses concerning means including the setting of confidence limits on means, and (3) in tests of hypotheses concerning scale parameters including the setting of confidence limits on scale parameters. Tukey (1960) and Cox (1954) have discussed various uses of scale parameters, estimators and other related topics.

For the first two uses outlined above, the relative biases of estimators may be important considerations, particularly in large samples. However, for the third use the relative biases of estimators become unimportant in tests for homogeneity of variance as long as all samples are of the same size and the parent distributions involved can be assumed to have essentially the same shapes. Tukey (1960) and Cox (1954) have also given a more detailed discussion of this point.

In Chapters 5, 6, and 7 we will be primarily interested in tests for homogeneity of variance and under the above assumptions the biases of the estimators will be of no importance. However, in the present chapter, information will also be accumulated on the expectation or bias of the estimators since it is readily available as a by-product in computation of the more pertinent coefficient of variation. It should be noted that only the bias of the estimators considered as estimators of σ, is obtained. The latter is an important point and will be discussed in more detail below.

Our primary purpose here is, however, to gain information about the sampling variation of the estimators, since Box and Anderson (1955, page 16) have shown that the reason tests on means are robust while tests on variance are not, is

> "... because, whereas in tests to compare means we compare the variation among the means with an estimate of the variation obtained from internal evidence within the groups, in current tests to compare variances ...we tacitly compare some measure of variation among the variances with a theoretical value which is correct only for the normal distribution."

A natural measure of variation in this context is the variance of the unbiased version of the estimator. To avoid pathological considerations we assume throughout that the expectation of an estimator is a constant multiple of σ, where the constant may depend on the parent population and on the sample size but not on σ. Then the variance of the unbiased version of the estimator is identical to its squared coefficient of variation defined by

$$CV^2(T) = V(T)/E^2(T) = V(T/E(T)) .$$

This measure of variation permits us to avoid the troublesome question of exactly what is being estimated.

We note here that it is not necessary to compute the expectation of the sample range for both samples of size two and three since

$$E(R_2) \equiv \tfrac{2}{3} E(R_3)$$

for any distribution. This may be seen by using Tippett's (1925) formula for $E(R)$, or by using a more general recurrence relation for $E(R)$ due to Carlson (1958).

Robustness of Scale Parameters

Before beginning a discussion of the complicated problem of robustness of scale parameter estimators, it may be informative to consider the related concept of robustness of scale parameters per se. Just as the various properties of a scale estimator assume varying importance, depending upon its usage, so do those of scale parameters. In scientific work scale parameters, or their estimators, are probably most frequently used in setting some sort of "uncertainty interval" about an estimated mean or other least squares value. One would hope that the frequency with which the interval covered the true value did not depend sensitively on the shape of the parent distribution. Thus, one natural measure of scale parameter robustness would seem to be the stability of the percent of the distribution covered by

$\mu \pm \theta$, where μ is the expected value and θ is a scaling parameter. In this framework σ is a different parameter from 3σ . Another quite similar measure of scale parameter robustness may be defined as the stability of the value of k needed to make $\mu \pm k\theta$ contain a specified probability.

Investigations by Eisenhart (unpublished), Chand (1949), Walsh (1956) and Pearson and Tukey (1965) gives preliminary evidence that for a large number of "reasonable" continuous distributions $k\sigma$ is a very robust scaling parameter if (roughly) $1\frac{3}{4} \leq k \leq 2$. Apparently no such investigations have been made for δ, the population mean deviation, or for the expected value of R. It would seem that such investigations might be fruitful in either pointing out the inherent non-robustness of the parent mean deviation and expected range or putting them on the firm footing presently enjoyed only by the standard deviation. A brief analysis of the robustness of the expected value of R in this context is included in Section 4.4 for Tukey random variables.

It may be noted that in the above sense the interpercentile distances (see Tukey 1960 for discussion) satisfy this criterion of robustness perfectly.

In the analysis below, a subtle bias is introduced by considering S, M, and R as estimators of σ . This assumption would seem to prejudice the case against M and R since $E(M)$ and $E(R)$ could be very robust scale parameters in the sense outlined above but not be very robust as estimators of σ . We will briefly compare the properties of M considered as an estimator of the parent mean deviation with its properties when considered as an estimator of σ for some discrete distributions in Section 4.5.

It may be re-emphasized that such considerations of expectation are irrelevant in tests for homogeneity of variance when samples are of the same size and are all from populations with the same shape, that is, populations that differ at most in location and scale.

4.2 PROPERTIES UNDER NORMALITY

The properties of S, M, and R have been extensively studied for samples from the normal distribution. The purpose of this section is to summarize such properties as are needed elsewhere in this work.

The Standard Deviation

The sample standard deviation S is defined by

$$S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / (n-1) .$$

For samples from a normal distribution it is well known that $\nu S^2/\sigma^2$ is distributed as chi-square, where $\nu = n-1$ is the degrees of freedom associated with S^2, and σ^2 is the variance of the parent distribution.

Pearson and Hartley (1954) have tabulated the expected value, standard deviation, β_1, β_2 and a number of other quantities for the standard deviation. The expected value of S/σ may be computed from the relation

$$\mu_\nu = E(S/\sigma) = \sqrt{2/\nu}\ \Gamma\left(\frac{\nu+1}{2}\right) / \Gamma\left(\frac{\nu}{2}\right) ,$$

where

$$\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy .$$

For general $r \geq 2$,

$$E\left[(S/\sigma)^r\right] = \frac{\nu + r - 2}{\nu} E\left[(S/\sigma)^{r-2}\right] .$$

The squared coefficient of variation of S based on ν degrees of freedom is given by

$$CV^2(S) = 1/\mu_\nu^2 - 1 .$$

The Mean Deviation

The sample mean deviation is defined by

$$M = \sum_{i=1}^{n} |X_i - \bar{X}|/n$$

A summary of known properties of the mean deviation in normal samples has recently been given by Herrey (1965). The population mean deviation δ is related to σ in a normal distribution by

$$\delta = \sqrt{2/\pi}\ \sigma .$$

For normal samples the

$$E(M/\delta) = \sqrt{(n-1)/n}$$

and

$$CV^2(M) = \frac{1}{n}\left[\tfrac{1}{2}\pi + \arcsin(n-1)^{-1} - n + \sqrt{n(n-2)}\ \right].$$

Godwin and Hartley (1945) have tabulated percentage points of M and Cadwell (1953, 1954a) has given fractional powered chi-square approximations to its distribution.

For samples of size three from a normal distribution the mean deviation and range have identical distributions (except for a scale factor). This result which does not appear to have been noticed before is stated in the following theorem.

<u>THEOREM</u> : For samples of size three from a normal distribution R has the same distribution as $(3\sqrt{3}\ /\ 2)M$.

PROOF :

$$\begin{aligned} 3M &= |X_1-\bar{X}| + |X_2-\bar{X}| + |X_3-\bar{X}| \\ &= \left|\frac{2X_1-X_2-X_3}{3}\right| + \left|\frac{2X_2-X_1-X_3}{3}\right| + \left|\frac{2X_3-X_1-X_2}{3}\right| \\ &= \sqrt{2/3}\ \left(|U| + |V| + |W| \right) \end{aligned}$$

where U, V, and W are trivariate normal with mean zero and covariance matrix

$$\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix} .$$

Similarly

$$\begin{aligned} 2R &= |X_1-X_2| + |X_2-X_3| + |X_3-X_1| \\ &= \sqrt{2}\ \left(|U'| + |V'| + |W'| \right) \end{aligned}$$

where U′, V′, and W′ have the same distribution as U, V, and W above.

This common distribution of M and R may be written as (McKay and Pearson, 1933)

$$f(y) = 6/\sqrt{2}\left(2\Phi(y/\sqrt{6}) - 1\right)\varphi(y/\sqrt{2})$$

where

$$Y = R = \frac{3\sqrt{3}}{2} M$$

and Φ and φ are the normal c.d.f. and p.d.f. respectively.

The Sample Range

The sample range is defined by

$$R = X_{max} - X_{min} .$$

An excellent summary of properties of the range for samples from a normal distribution has recently been given by David (1962). Extensive tables of the probability integral, percentage points, and moments of the range have been given by Harter and Clemm (1959) and Harter (1960). Percentage points of the ratio of two ranges have been tabulated by Harter (1963).

Relative Efficiencies

We will be concerned in Chapter 7 below, with the properties of

1. the pooled standard deviation S_p defined by

$$S_p^2 = \Sigma S_j^2 /k,$$

2. the average standard deviation

$$\bar{S} = \Sigma S_j/k,$$

3. the average mean deviation

$$\bar{M} = \Sigma M_j/k,$$

and 4. the average range

$$\bar{R} = \Sigma R_j/k,$$

where in each definition the sums are from $j=1$ to k and the k estimates (S_j, M_j, or R_j, as appropriate) are assumed to be identically and independently distributed. In particular, each of the k samples has the same number of observations.

For samples from a normal distribution, S_p has the same distribution as an individual standard deviation based on $k\nu$ degrees of freedom. For S_p the

$$CV^2(S_p) = 1/\mu_{k\nu}^2 - 1 \quad .$$

The coefficient of variation of an average is related to the coefficient of variation of an individual observation by

$$CV^2(\bar{Y}) = CV^2(Y)/k \; .$$

Hence

$$CV^2(\bar{S}) = CV^2(S)/k,$$

etc.

Let the relative efficiency of an estimator Z with respect to an estimator Y be defined by the ratio of the variances of the unbiased estimators. That is, let

$$\mathrm{Eff}(Z|Y) = V(Y/E(Y))/\, V(Z/E(Z))$$

which is equivalent to

$$\mathrm{Eff}(Z|Y) = CV^2(Y) \;/\; CV^2(Z) \; .$$

Then Table 4.2.1 gives the relative efficiency of $\bar{S}$ with respect to S_p . Also given is the relative efficiency of M with respect to S and the relative efficiency of R with respect to S. The relative efficiency of $\bar{M}$ or $\bar{R}$ with respect to $\bar{S}$ for any given n, is the same for all k. The relative efficiency of $\bar{M}$ with respect to S_p may readily be computed as

$$\mathrm{Eff}(\bar{M}|S_p) = \mathrm{Eff}(M|S) \cdot \mathrm{Eff}(\bar{S}|S_p) \; .$$

An analogous relation holds for the efficiency of $\bar{R}$ with respect to S_p .

For samples from a normal distribution, the properties of S for a sample of size n in which the mean must be estimated, are the same as those for a sample of size n-1 where the mean is known. This is not true for the mean deviation although its expectation does behave in an analogous manner.

Table 4.2.1

(a) Efficiency of $\bar{S}$ with respect to S_p. Total of k estimates each based on n observations from a normal distribution.

k \ n =	2	3	4	5	6	8	10	12	15	20	50	∞
1.	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2.	.957	.964	.971	.976	.980	.985	.988	.990	.992	.994	.998	1.000
3.	.936	.950	.960	.968	.973	.979	.983	.986	.989	.992	.997	1.000
4.	.923	.942	.954	.963	.969	.977	.981	.984	.987	.991	.996	1.000
5.	.915	.937	.951	.960	.967	.975	.980	.983	.987	.990	.996	1.000
6.	.909	.933	.948	.958	.965	.974	.979	.982	.986	.990	.996	1.000
8.	.902	.929	.945	.956	.963	.972	.978	.982	.985	.989	.996	1.000
10.	.897	.926	.943	.954	.962	.971	.977	.981	.985	.989	.995	1.000
15.	.890	.922	.941	.953	.960	.970	.976	.980	.984	.988	.995	1.000
25.	.885	.919	.939	.951	.959	.969	.976	.980	.984	.988	.995	1.000
50.	.880	.917	.937	.950	.958	.969	.975	.979	.983	.988	.995	1.000
100.	.878	.916	.937	.949	.958	.968	.975	.979	.983	.987	.995	1.000
∞	.876	.915	.936	.949	.957	.968	.975	.979	.984	.987	.995	

(b) Efficiency of M with respect to S

n =	2	3	4	5	6	8	10	12	15	20	50	∞
	1.000	.992	.964	.946	.934	.919	.910	.904	.898	.893	.883	.876

(c) Efficiency of R with respect to S

n =	2	3	4	5	6	8	10	12	15	20	50	∞
	1.000	.992	.975	.955	.933	.890	.850	.814	.766	.700	.488	.000

4.3 A SURVEY OF INFORMATION AVAILABLE ON THE FIRST TWO MOMENTS OF S, M, AND R FOR SAMPLES FROM NON-NORMAL POPULATIONS

Analytic Results on S

Small sample analytic approaches to the properties of S from non-normal distributions have been almost uniformly unsuccessful. Such meagre results as are available are outlined below in chronological order. Results for sample size n = 2 are generally not included since they are more appropriately considered as results on the range.

Rietz (1931)	Uniform (0,1) parent; exact distribution for n = 3.
Craig (1932)	Integral equations for joint distribution of $\bar{X}$ and S from arbitraty parent for n = 3,4. Joint distribution given explicitly for exponential and uniform for n = 3.
Baker (1935)	Two-term Gram-Charlier parent; joint distribution of $\bar{X}$ and S for n = 2,3. Three-term Gram-Charlier parent; joint distribution of $\bar{X}$ and S for n = 2.
Truksa (1940)	Two-fold integral recurrence relationship for joint cumulative distribution of $\bar{X}$ and S.
Irick (1950)	(n-1)-fold integral representation for distribution of S.

Springer (1950) Distribution of S from χ_2^2 for arbitrary n and from χ_4^2 for n = 3. (Results seem to be in error since $E(S^2)$ does not check.)

Springer (1953) Integral recurrence relationship for joint distribution of $\bar{X}$ and S from doubly infinite parent.

Bennett (1955) Integral recurrence relationship for joint distribution of $\bar{X}$ and S.

Sampling Studies on S

E. S. Pearson and others performed a sampling study for six Pearson curves about 1928. Results on the sample variance obtained in this study for samples of size 5, 10, and 20 are given by Le Roux (1931). Also included in Le Roux's tables are results of a sampling study he performed for n = 5,25 and results of a study by Church (1926) for n = 10. The grouping interval of the reported results on the variance was very fine (approximately 50 categories) so the expectation and variance of the standard deviation could be satisfactorily computed from the reported results. Table 4.3.1 gives the results of this calculation.

Neyman and Pearson (1928) report, for Church's population, $E(S) = 3.3181$ and $\sigma_S = 0.8086$, whereas, from Le Roux's table, I get $E(S) = 3.3096$ and $\sigma_S = 0.8096$. This slight discrepancy is probably due to their having additional information but is certainly not large enough to change any of our conclusions.

Table 4.3.1

Results on S computed from sampling studies reported by Le Roux (1931). "Standardized" means that the entry has been divided by the normal theory equivalent. The "*" indicates that the result was for n = 25.

PEAR-SON TYPE	B2	DL/SG	STANDARDIZED EXPECTATION N=5	10	20	STANDARDIZED COEF. VAR.SQ N=5	10	20	HUNDREDS OF SAMPLES N=5	10	20
II	2.50	.8184	1.02	1.00	1.01	.86	.84	.77	10	5	5
VII	4.12	.7714	1.00	.99	1.01	1.31	1.19	1.45	10	5	5
VII	7.05	.7369	.96	.96	.98	1.73	2.17	2.27	10	5	5
III	3.29		1.01	1.01	1.00	1.21	1.03	1.10	10	5	5
III	3.72	.7909	.99	.98	.96	1.15	1.17	1.55	10	5	10
I	3.83	.7971	.98	.99	.98	1.29	1.45	1.45	10	10	5
IV	5.81		.95	.95	1.01*	1.37	1.05	2.08*	5	10	5*

Neyman and Pearson (1928) used beads to do a sampling study on the standard deviation for samples from a discrete distribution that was approximately rectangular. For samples of size n = 4 they used a population with K = 21 categories and for n = 10, K = 11 categories.

Dunlap (1931) performed a sampling study on the standard deviation using dice. The discrete parent distribution was approximately rectangular with K = 6 categories and samples of size 10 were taken.

Bounds and Asymptotic Results on S

The first six even moments of S about the origin are available in terms of the lower parent cumulants since they are simply the integral moments of S^2 (see, for example, Kendall and Stuart, 1958). Asymptotic expansions for the odd moments of S in terms of the parent moments have been given by Craig (1929) and Kondo (1930) but for small sample sizes these expansions seem to converge only for distributions very near the normal (see Pearson 1929).

Le Roux (1931) has made the interesting discovery that there exists a subset of the Pearson family of distributions (corresponding to a curved line in the Type I region) for which the sample variance based on $\nu = n-1$ degrees of freedom is very nearly distributed as the sample variance of a sample based on ν^* degrees of freedom from a normal distribution where

$$\nu^* = \nu \left[1 + \frac{1}{2}\frac{\nu}{\nu+1}(\beta_2 - 3)\right]^{-1} .$$

He has given $\beta_1 - \beta_2$ diagram showing this subset on Page 149. This "equivalent degrees of freedom" procedure gives the correct value for $CV^2(S^2)$ for all parent populations and it will be shown below in Section 4.5 that it seems to provide a good working approximation to the behavior of S for samples from some discrete distributions.

Harris and Tukey (1949) give asymptotic efficiencies of the standard deviation and mean deviation for samples from "contaminated normal" distributions for various amounts of "contamination." In their contaminated normal model each observation has (independent) probability P of coming from a normal distribution with standard deviation 3σ, and probability (1-P) of coming from a normal distribution with standard deviation 1σ, both with mean μ. Table 4.3.2 gives the asymptotic relative efficiency of M with respect to S computed from their results. P denotes percent contamination and asymptotic efficiency is defined to be $CV^2(S)/CV^2(M)$.

Table 4.3.2

Asymptotic efficiency of M relative to S for contaminated distributions

P	β_2	$CV^2(S)/CV^2(M)$
0	3.00	0.86
1	4.63	1.44
2	5.80	1.75
5	7.65	2.03
10	8.33	1.90

The following bounds on $E(S/\sigma)$ may be found using Liapunov's inequality (Kendall and Stuart, 1958, page 63):

$$\left[\frac{n(n-1)}{(n-1)(\beta_2-3) + n(n+1)} \right]^{\frac{1}{2}} \leq E(S/\sigma) \leq 1$$

Thus bounds on the coefficient of variation of S are given by

$$0 \leq CV^2 (S) \leq \frac{2}{n-1} + \frac{(\beta_2-3)}{n} .$$

Results on S^2

The lower moments of S^2/σ^2 are well known (see for example Kendall and Stuart, 1958, p.290).

$$E(S^2/\sigma^2) \equiv 1$$

$$CV^2(S^2) = \frac{(\beta_2-3)}{n} + \frac{2}{n-1}$$

Results on M

The properties of the mean deviation in samples drawn from non-normal populations are even less well known than those of the standard deviation. The expectation and approximate variance of M in samples from a Pearson Type III distribution were given by Johnson (1958), but no other analytical work on the mean deviation seems to have succeeded.

Asymptotically the variance of M is given by

$$V(M) = \left(\sigma^2 - \delta^2\right) / n + 0\left(n^{-2} \right)$$

where $\delta = E|X-\mu|$ (see Kendall, 1950, exercise #194).

To the author's knowledge no sampling studies have been reported on the mean deviation, though in sampling studies of Geary's (1935) criterion for testing for non-normality, values of S and M must have been computed.

Results on R

The range, being more mathematically tractable, has been subjected to a much more thorough investigation. Pearson (1950), David (1954), and Cox (1954) have summarized much of what was then known about the small sample properties of the range. In addition, the exact distribution of R has been given for samples from a discrete rectangular population by Rider (1951) and for samples from the logistic distribution by Gupta (1965). Burr (1966) has recently given the expectation and coefficient of variation of R for samples from a family of distributions defined by

$$F(X) = 1 - \left(1 + X^c \right)^{-k} \qquad 0 \leq X < \infty .$$

Asymptotic results on R have been given (e.g., by Cadwell, 1954) but are not included here since our primary interest is in small sample properties and a number of small sample results are available for the range.

Burr (1955) has given a computational formula for the distribution of R for samples from discrete distributions, and has computed the distribution and first four moments of R for an 11-category analog of the normal and a 12-category analog of a Γ distribution. More results on discrete distributions will be given in Section 4.5.

In addition to the sampling studies reported by Pearson (1950), Gephart (1955) has obtained empirical results for 13 populations for samples of size 3, 5, 7 and 10. The relevant part of his results is outlined in Table 4.3.3 below.

Tables 4.3.4, 4.3.5, and 4.3.6 outline the numerical results on R available in the literature. The column headings are abbreviations for the following:

$$B2 = \mu_4/\mu_2^2 \quad \text{and} \quad DL/SG = \delta/\sigma = E|X-\mu| / \sigma$$

for parent population.

Under -RANGE-,

D means distribution of range is given,

E means expectation of range is given,

V means variance of range is given.

Under ORDER STAT,

E means expectations of the order statistics are given,

V means covariances of the order statistics are given.

Entries in the columns denote the following:

T means numerical results are given in tables,

F means formulas only are given,

B means both formulas and tables are given.

Table 4.3.4 lists the populations for which information on R is available. Table 4.3.5 gives, for these populations, the standardized expectation of R defined by $E(R/d_n\sigma)$ where $E(R/\sigma) = d_n$ for samples from the normal distribution. Table 4.3.6 gives the standardized coefficient of variation of R which is defined by $CV^2(R)/[CV^2(R)$ from normal distribution].

The mixed normal distributions referred to in Tables 4.3.4, 4.3.5 and 4.3.6, through the ID of MNk, are defined such that each observation has (independently) a 50% chance of coming from a normal distribution with mean $-k\sigma$ and a 50% chance of coming from a normal distribution with mean $+k\sigma$.

In Table 4.3.4 the reference entry is sometimes abbreviated and occasionally a second or third author's name has been omitted entirely.

Hartley and David (1954) have given a lower bound for $E(R/\sigma)$ for distribution on the finite interval $a \leq x \leq b$.

Table 4.3.3

Empirical results due to Gephart (1955). Each entry based on 2000 samples. (Two sigma limits on $E(R/d_n\sigma)$ are roughly $\pm .05$). Some exact results are available on four of his populations and are listed in Tables 4.3.5 and 4.3.6.

			Expectation				Coeff. Var. Sq.			
ID	B2	DL/SG	N= 3	5	7	10	N= 3	5	7	10
NORMAL	3.00	.7979	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
EDGWTH	4.50		.986	1.001	.997	1.001	1.057	1.035	1.043	1.033
EDGWTH	6.00		.954	.988	.997	1.022	1.326	1.515	1.642	1.744
EDGWTH	6.75		.926	.943	.979	1.019	1.779	2.101	2.420	2.561
EDGWTH	4.50		.868	.933	.951	1.003	2.024	2.501	2.791	3.071
EDGWTH	6.00		.989	.969	1.004	1.025	1.230	1.596	1.589	1.706
EDGWTH	4.50		.932	.967	.990	1.031	1.696	2.128	2.243	2.548
EDGWTH	6.00		.957	.973	.988	.993	1.386	1.546	1.543	1.779
EDGWTH	5.40		.924	.935	.952	.975	1.595	1.938	2.185	2.446
GM 16	3.38	.7937	.980	.965	.975	.986	1.363	1.588	1.637	1.749
GM 4	4.50	.7815	.993	1.005	1.001	1.000	1.094	1.167	1.176	1.227
GM16/9	6.38	.7617	.973	.971	.977	.977	1.265	1.344	1.444	1.501
EXP	9.00	.7358	.928	.944	.944	.952	1.605	1.769	1.982	2.156

Table 4.3.4

List of distributions for which information on the first two moments of R is available.

For explanation of notation see text.

ID	POPULATION	B2	DL/SG	RANGE D	RANGE E	RANGE V	ORDER STAT. F	ORDER STAT. V	REFERENCE
N	NORMAL	3.00	.7979	T	T	T			HARTER 1960
R	RECTANGULAR	1.80	.8660	F	B	B			MCKAY PEARSON 1933
ST	SYMMET TRIANG	2.40	.8165				T	T	SARHAN 1954
RT	RIGHT TRIANG	2.40	.8381	F	B	B			MCKAY 33 OSTLE 61
CHI-1	CHI 1-DEG FDM	3.87	.8006				B	B	GOVIND. 65 FOLKS 65
LG	LOGISTIC	4.20	.7642	B	B			B	BIRNB. 63, SHAH 65
EV	EXTREME VALUE	5.40		F	B			F	DAVID 54,LIEBLEIN 53
DE	DBL EXPONENTAL	6.00	.7071		T		B	B	MOORE 57,GOVIND. 66
E	EXPONENTIAL	9.00	.7358	B	B	B			DAVID 1954 COX 1954
GM2	GAMMA A=2	6.00	.7656	F	B	T			DAVID 1954
GM3	GAMMA A=3	5.00	.7761		T		B		MOORE 57, GUPTA 60
GM4	GAMMA A=4	4.50	.7815		T		B		MOORE 57, GUPTA 60
GM5	GAMMA A=5	4.20	.7847				B		GUPTA 1960
GM9	GAMMA A=9	3.67	.7905		T	T			MOSES 1956
GM16	GAMMA A=16	3.38	.7937		T	T			MOSES 1956
W1.5	WEIBULL A=1.5	4.39	.7878				B		HARTER 1964
W2	WEIBULL A=2	3.25	.8038				B		HARTER 1964
W3.5	WEIBULL A=3.5	2.71	.8084				B		HARTER 1964
W6	WEIBULL A=6	3.04	.7996				B		HARTER 1964
W8	WEIBULL A=8	3.33	.7939				B		HARTER 1964
T5	STUDENT NU=5	9.00	.7351		T	T			MOSES 1956
T8	STUDENT NU=8	4.50	.7655		T	T			MOSES 1956
T13	STUDENT NU=13	3.67	.7800		T	T			MOSES 1956
B3	BETA A=B=3	2.33	.8267		T	T			MOSES 1956
B10	BETA A=B=10	2.74	.7969		T	T			MOSES 1956
MN0.5	MXD NORMAL *	2.92	.8010		T	T			MOSES 1956
MN1.0	MXD NORMAL *	2.50	.8249		T	T			MOSES 1956
MN1.5	MXD NORMAL *	2.04	.8646		T	T			MOSES 1956
MAX	E(R/SIGMA) ALL POPULATIONS				B				PLACKETT 1947
MIN	CV2(R) ALL POPULATIONS					B			MORIGUTI 1954

Table 4.3.5

Summary of results available on $E(R/d_n\sigma)$.

Standardized Expectation of the Range

ID	B2	n = 2,3	4	5	6	8	10
N	3.00	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
R	1.80	1.02333	1.00957	.99289	.97630	.94630	.92096
ST	2.40	1.01304	1.00944	1.00506	1.00055	.99172	.98344
RT	2.40	1.00265	.99441	.98442	.97442	.95611	.94037
CHI-1	3.87	.97176	.96907	.96580	.96255	.95667	.95171
LG	4.20	.97720	.98192	.98765	.99342	1.00415	1.01361
EV	5.40	.95791	.96083	.96437	.96793	.97456	.98039
DE	6.00	.93999	.95168	.96587	.98000	1.00564	1.02743
E	9.00	.88623	.89051	.89570	.90093	.91067	.91924
GM2	6.00	.93999	.94194	.94431	.94671	.95117	.95512
GM3	5.00	.95937	.96062	.96214	.96367	.96653	.96906
GM4	4.50	.96931	.97023	.97135	.97247	.97457	.97643
GM5	4.20	.97535	.97608	.97696	.97784	.97949	.98095
GM9	3.67			.98708			
GM16	3.38			.99270			
W1.5	4.39	.96600	.96545	.96481	.96415	.96299	.96203
W2	3.25	.99314	.99128	.98899	.98671	.98249	.97883
W3.5	2.71	1.00627	1.00458	1.00249	1.00040	.99647	.99300
W6	3.04	.99793	.99752	.99697	.99641	.99536	.99437
W8	3.33	.99152	.99173	.99200	.99226	.99272	.99306
T5	9.00			.96639			
T8	4.50			.98618			
T13	3.67			.99344			
B3	2.33			.99790			
B10	2.74			1.00227			
MN0.5	2.92			1.00068			
MN1.0	2.50			1.00187			
MN1.5	2.04			.99509			
MAX	ALL	1.02333	1.01224	1.00610	1.00746	1.02584	1.05423

Table 4.3.6

Summary of information available on

$CV^2(R_n)$ / [$CV^2(R_n)$ from normal distribution].

Squared Coefficient of Variation of the Range

ID	B2	n = 2	3	4	5	6	8	10
N	3.00	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
R	1.80	.8760	.7260	.6084	.5176	.4466	.3446	.2761
ST	2.40		.8766	.8310	.7900			
RT	2.40	.9854	.9075	.8402	.7835	.7356	.6594	.6011
CHI-1	3.87	1.1623	1.1859	1.2022	1.2155	1.2271	1.2474	1.2645
LG	4.20	1.1299	1.2100	1.2769	1.3332	1.3813	1.4598	1.5219
EV	5.40							
DE	6.00							
E	9.00	1.7519	2.0167	2.2174	2.3766	2.5073	2.7122	2.8683
GM2	6.00							
GM3	5.00							
GM4	4.50							
GM5	4.20							
GM9	3.67				1.1435			
GM16	3.38				1.0803			
W1.5	4.39							
W2	3.25							
W3.5	2.71							
W6	3.04							
W8	3.33							
T5	9.00				1.7017			
T8	4.50				1.3439			
T13	3.67				1.1808			
B3	2.33				.8515			
B10	2.74				.9128			
MN0.5	2.92				.7984			
MN1.0	2.50				.8114			
MN1.5	2.04				.6272			
MIN	ALL	.8760	.6907	.498		.207	.074	

Comparison of S and R

Pearson and Haines (1935) plot S vs R for some real data for n = 5,6, and 10. They also compare some sampling results for three populations with β_2 = 2.5, 4.1 and 3.3 and find for n = 5 and 10 that there is no significant difference between S and R in the number of points falling above the normal theory 5 and .5 percent points.

David (1954) compares the probability that S and R fall beyond the normal theory 5% points when the parent distribution is log χ^2.

Cox (1954) compares the standard errors of estimate and the coefficient of variation of the average range and the pooled standard deviation, and finds, for sufficiently large numbers of samples that for

> small and moderate values of β_2 the root-mean-square estimate is the better one, but that for larger values of β_2 the estimate based on the range, even though biased, is to be preferred. The reason is essentially that for populations with large β_2 very extreme observations are common, and these have relatively less effect on the range than the other estimate.

Further comparisons are now possible and will be discussed in more detail below.

4.4 PROPERTIES OF THE RANGE IN SAMPLES FROM SYMMETRIC TUKEY RANDOM VARIABLES

In this section we derive some properties of the sample range for samples from a family of random variables introduced by Tukey (see Hastings, et al., 1947). First we give some general properties of this family of random variables, then derive closed form expressions for the first two moments of R. We then compute the expected value and coefficient of variation of R for a number of members of the family. These results are compared with those available from other distributions in Section 4.6. We conclude this section with a discussion of the "robustness" of $E(R_n)$ as a scale parameter in the sense described in Section 4.1. This family has also been discussed briefly by Tukey (1962).

Introduction

Let

(4.4.1) $$Z = \frac{1}{\lambda}\left(U^{\lambda} - (1-U)^{\lambda}\right) \quad \lambda \neq 0$$

where U is uniformly distributed on $(0,1)$. Then Z is said to be distributed as a symmetric Tukey random variable. A limiting form of the transformation (4.4.1) as $\lambda \to 0$ is[1] $Z = \log \frac{U}{1-U}$, and then Z is distributed according to the logistic distribution.

---

[1] This can be seen by noting that

$$\lim_{\lambda\to 0} \frac{1}{\lambda}\left(u^{\lambda} - (1-u)^{\lambda}\right) = \lim_{\lambda\to 0} \int_{1-u}^{u} t^{\lambda-1}dt = \log \frac{u}{1-u} .$$

A location and scale parameter can be added to the transformation if desired, but are unnecessary for present purposes. The family can be generalized to the non-symmetric case

$$Z = AU^{\lambda} - B(1-U)^{\gamma} + C$$

but we restrict attention to the symmetric case outlined above. The non-symmetric family $Z = AU^{\lambda} - (1-U)^{\lambda}$ has been used by Hogben (1963) and Shapiro and Wilk (1965) in sampling studies.

Parent Properties

The range of variation of Z is

$$-\frac{1}{\lambda} \leq Z \leq \frac{1}{\lambda} \qquad \lambda > 0$$

$$-\infty \leq Z \leq \infty \qquad \lambda \leq 0 .$$

The density function of Z is defined implicitly by

$$g(z(u)) = \left(u^{\lambda-1} + (1-u)^{\lambda-1} \right)^{-1}$$

and the ordinates at the extremes of the range of variation of Z are given by

$$\begin{aligned} g(z(0)) = g(z(1)) &= 1 \qquad \lambda > 1 \\ &= \tfrac{1}{2} \qquad \lambda = 1 \\ &= 0 \qquad \lambda < 1 . \end{aligned}$$

When $\lambda=1$ or 2, the transformation is linear so the distribution of Z is uniform. The density of Z is slightly U-shaped for $1 < \lambda < 2$ having a minimum value of $\beta_2 = \mu_4 / \mu_2^2$ of approximately 1.75 at approximately $\lambda = 1.45$. The density has a single mode for $\lambda < 1$ or $\lambda > 2$.

Tukey (1960a) and Van Dyke (1961) have found that this family of distributions can be used to give useful approximations to the percentage points of the normal, and t distributions. Some results based on their approximations are given below.

An important property of Tukey random variables is that the percentage points are available directly from the definition. That is, the 100 p'th percent point of Z is given by $z(p)$ where $z(u)$ is defined in (4.4.1).

In the remainder of this discussion, the case $\lambda=0$ for which Z has the logistic distribution is omitted.

The odd moments of Z (which exist) are zero by symmetry and the even moments are

$$E(Z^k) = \frac{2}{\lambda^k} \sum_{j=0}^{m-1} (-1)^j \binom{k}{j} B(j\lambda + 1, \lambda(k-j) + 1)$$

$$+ (-1)^m \binom{k}{m} B(m\lambda + 1, m\lambda + 1)$$

where $m = k/2$

and

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}dx \; .$$

Note that the moment of order k does not exist if $\lambda \leq -1/k$.

The absolute moments are given by

$$E|Z^k| = \frac{2}{\lambda^k} \sum_{j=0}^{m} (-1)^j \binom{k}{j} \left[2B_{\frac{1}{2}}(a,b) - B(a,b) \right]$$

where

$$2m + 1 = k, \quad a = j\lambda + 1, \quad b = (k-j)\lambda + 1$$

and

$$B_{\frac{1}{2}}(p,q) = \int_0^{\frac{1}{2}} x^{p-1}(1-x)^{q-1}dx \quad .$$

The formulas above simplify for the lower moments to

$$\delta = E|Z| = \frac{2}{\lambda(\lambda+1)} \left[1 - (\tfrac{1}{2})^\lambda \right]$$

$$\sigma^2 = \frac{2}{\lambda^2(2\lambda+1)} \left[1 - \tfrac{1}{2}\lambda B(\lambda,\lambda) \right]$$

$$\mu_4 = \frac{2}{\lambda^4(4\lambda+1)} \left[1 - 3\lambda B(3\lambda,\lambda) + 3\lambda B(2\lambda,2\lambda) \right] \quad .$$

For large λ the asymptotic formula

$$\beta_2 \approx \tfrac{1}{2}(\lambda + \tfrac{3}{4})$$

may be derived from the formulas above and for $-.23 \leq \lambda \leq -.03$ the emperical formula

$$\beta_2 \doteq (\lambda + .246)^{-1}$$

is accurate to within about 2 percent.

The flexibility of the symmetric TRV family is demonstrated by the inclusion of such diverse distributions as :

1. Rectangular
2. Logistic
3. A very good approximation to the normal
4. Very good approximations to t-distributions
5. Distributions for which Plackett's upper bound on $E(R_n/\sigma)$ is attained. (This will be shown below.)

<u>Moments of R</u>

The ℓ'th order statistic in a sample of size n is distributed as $Z_{n,\ell} = Z(U_{n,\ell})$ where $U_{n,\ell}$ is the ℓ'th order statistic in a sample of size n from the uniform distribution. Hence, percentage points of $Z_{n,\ell}$ are available by using the defining relation (4.4.1) in conjunction with tables of the incomplete beta function or tables of percentage points of order statistics from a uniform distribution (see Govindarajulu and Hubacker, 1964).

The moments of the order statistics are

$$E(Z_{n,\ell}^{k}) = \frac{\ell}{\lambda^k} \binom{n}{\ell} \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} B(j\lambda+\ell,\ \lambda(k-j)+n-\ell+1) .$$

The product moment of smallest and largest is

$$\frac{\lambda^2}{n(n-1)} \cdot E(Z_{n,1}\, Z_{n,n}) = \frac{2}{2\lambda+n} B(\lambda+1,n-1) + \Big((-1)^n - 1\Big) B(\lambda+n,\lambda+1)$$

$$\cdot\ B(\lambda+1,n-1) - \sum_{j=0}^{n-2} (-1)^j \binom{n-2}{j} \frac{B(\lambda+1,\ j+1)}{n+\lambda-j-1} \ .$$

Other formulas for the expectation, variance and covariance of the order statistics were given by Hastings, et al. (1947).

We are particularly interested in the sample range R, its expectation $E(R)$ and its squared coefficient of variation $CV^2(R)$ which may readily be obtained from the expressions (4.4.2) and (4.4.3). In particular

$$E(R) = \frac{2n}{\lambda(n+\lambda)} \Big[1 - \lambda B(n,\lambda)\Big] \ .$$

Plackett (1947) has given an upper bound on the ratio $E(R_n/\sigma)$ that holds for all parent distributions. He notes that this upper bound is attained for a sample of size n if the distribution function of X is defined implicitly by $X = F^{n-1} - (1-F)^{n-1}$, where F is the cumulative distribution function of the random variable X. Thus $E(R_n/\sigma)$ attains its upper bound for a symmetric Tukey random variable with $\lambda = n-1$.

Results of Computations

Let d_n denote the value of $E(R_n / \sigma)$ for samples from a normal distribution. Then Table 4.4.1 gives computed values of $E_n = E(R_n / d_n\sigma)$. Results are given for $n = 3(1)6(2)12,15,20$ and $\beta_2 = 2,3,4.2,6,8,10,15$. It is not necessary to give E_2 since, as has been pointed out in Section 4.1,

$$E(R_2 / d_2\sigma) \equiv E(R_3 / d_3\sigma)$$

for all distributions. Two entries are given for each (β_2,n) combination since there are two values of λ corresponding to any given value of β_2. The upper entry corresponds to the smaller value of λ. The upper entries for $\beta_2 = 4.20$ correspond to $\lambda=0$ (the logistic distribution) and were computed using the relation

$$E(R_n) = 2\sum_{j=1}^{n} 1/j \; .$$

The values of λ that have been used are indicated below :

β_2	2	3	4.2	6	8	10	15
Small	.585	.135	0	-.0802	-.1223	-.1466	-.17875
Large	2.82	5.20	7.64	11.25	15.25	19.25	29.25

Table 4.4.2 gives the corresponding values of $C_n = CV^2(R_n) / [CV^2(R_n)$ from normal distribution]. The upper entries for $\beta_2 = 4.20$ were obtained from the covariances of order statistics for the logistic distribution tabulated by Shah (1966) and Gupta, Qureishi, and Shah (1967).

Table 4.4.1

$E(R_n/d_n\sigma)$ for samples from symmetric Tukey random variables, where $d_n = E(R_n/\sigma)$ for normal distribution. Upper entry is for the smaller value of λ and lower for the larger value of λ.

B2=	2	3	4.2	6	8	10	15
2,3	1.0217	.9997	.9772	.9551	.9394	.9287	.9123
	1.0216	.9874	.9278	.8422	.7654	.7048	.5983
4	1.0119	.9999	.9819	.9628	.9488	.9390	.9239
	1.0121	.9934	.9461	.8712	.8002	.7423	.6373
5	1.0001	1.0001	.9876	.9722	.9601	.9515	.9579
	1.0006	1.0007	.9682	.9063	.8424	.7878	.6847
6	.9883	1.0004	.9934	.9816	.9716	.9642	.9521
	.9890	1.0073	.9895	.9407	.8842	.8333	.7326
8	.9668	1.0008	1.0042	.9993	.9932	.9880	.9759
	.9674	1.0169	1.0250	1.0006	.9592	.9165	.8228
10	.9484	1.0011	1.0136	1.0151	1.0124	1.0093	1.0029
	.9484	1.0219	1.0511	1.0483	1.0214	.9674	.9031
12	.9326	1.0013	1.2019	1.0290	1.0295	1.0253	1.0244
	.9319	1.0238	1.0700	1.0859	1.0727	1.0474	.9739
15	.9128	1.0015	1.0327	1.0473	1.0520	1.0533	1. 529
	.9107	1.0228	1.0890	1.1284	1.1336	1.1212	1.0651
20	.8868	1.0018	1.0474	1.0724	1.0832	1.0882	1.0927
	.8823	1.0162	1.1055	1.1747	1.2057	1.2126	1.1861

Table 4.4.2

$C_n = CV^2(R_n) \;/\; [CV^2(R_n) \text{ from normal distribution}]$ for samples of size n from symmetric Tukey random variables. Upper entry is for the smaller value of λ, and lower for the larger value of λ.

n	β_2 = 2	3	4.2	6	8	10	15
2	.8846	1.0018	1.1299	1.2649	1.3663	1.4385	1.5544
	.8846	1.0708	1.4451	2.1277	2.9459	3.7888	5.9365
3	.7709	1.0037	1.2100	1.4158	1.5665	1.6726	1.8412
	.7720	1.1155	1.6690	2.6293	3.7653	4.9318	7.8999
4	.6805	1.0047	1.2769	1.5444	1.7391	1.8757	2.0923
	.6818	1.1284	1.7785	2.8708	4.1519	5.4658	8.8110
5	.6089	1.0050	1.3332	1.6547	1.8884	2.0523	2.3122
	.6085	1.1139	1.8102	2.9600	4.3043	5.6843	9.2062
6	.5512	1.0048	1.3813	1.7507	2.0193	2.2079	2.5070
	.5477	1.0821	1.7958	2.9648	4.3317	5.7375	9.3347
8	.4646	1.0034	1.4598	1.9109	2.2402	2.4718	2.8397
	.4528	.9952	1.7009	2.8545	4.2100	5.6105	9.2130
10	.4026	1.0013	1.5219	2.0411	2.4219	2.6903	3.1173
	.3823	.9021	1.5752	2.6833	3.9949	5.3573	8.8798
12	.3561	.9988		2.1506	2.5761	2.8767	3.3557
	.3283	.8148	1.4474	2.4998	3.7563	5.0683	8.4788
15	.3044	.9947	1.6350	2.2875	2.7710	3.1135	3.6605
	.2679	.7013	1.2721	2.2380	3.4066	4.6371	7.8621
20	.2466	.9878	1.7140	2.4676	3.0308	3.4313	4.0732
	.2006	.5561	1.0355	1.8688	2.8987	3.9984	6.9209

Table 4.4.3

Results for TRV approximations to t distributions

(a) $E(R_n/d_n\sigma)$

d.f.	β_2 True	β_2 TRV	λ	n = 2,3	4	5	6	8	10	12	15	20
∞	3.00	2.97	.140	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
∞	3.00	3.00	.135	1.000	1.000	1.000	1.000	1.001	1.001	1.001	1.002	1.002
30	3.23	3.22	.099	.995	.996	.998	1.000	1.003	1.005	1.007	1.010	1.014
25	3.29	3.27	.092	.994	.996	.997	.999	1.003	1.006	1.008	1.012	1.015
20	3.38	3.35	.081	.992	.994	.997	.999	1.003	1.007	1.010	1.014	1.020
15	3.55	3.50	.062	.989	.992	.995	.998	1.004	1.009	1.013	1.019	1.026
13	3.67	3.60	.051	.987	.990	.994	.997	1.004	1.010	1.015	1.021	1.030
10	4.00	3.88	.025	.983	.986	.991	.996	1.004	1.012	1.019	1.027	1.039
8	4.50	--	-0.003	.977	.981	.987	.993	1.004	1.014	1.022	1.033	1.048
5	9.00	6.32	-0.089	.952	.960	.970	.980	.998	1.015	1.029	1.048	1.075
4	∞	10.04	-0.147	.929	.939	.951	.964	.988	1.009	1.028	1.053	1.088
3	∞	45.69	-0.224	.884	.897	.913	.930	.960	.988	1.013	1.047	1.093

(b) $CV^2(R_n)$ / [$CV^2(R_n)$ from normal distribution]

d.f.	n = 2	3	4	5	6	8	10	12	15	20
∞	1.002	1.004	1.005	1.005	1.005	1.003	1.001	.999	.995	.988
∞	.999	.998	.997	.996	.995	.991	.987	.984	.978	.969
30	1.028	1.047	1.062	1.075	1.085	1.100	1.111	1.119	1.127	1.137
25	1.033	1.056	1.075	1.089	1.102	1.120	1.134	1.144	1.156	1.169
20	1.043	1.071	1.095	1.114	1.129	1.154	1.172	1.186	1.202	1.221
15	1.060	1.099	1.132	1.158	1.180	1.216	1.243	1.264	1.289	1.319
13	1.071	1.117	1.154	1.186	1.212	1.254	1.286	1.312	1.343	1.380
10	1.099	1.161	1.213	1.257	1.293	1.353	1.399	1.437	1.482	1.539
8	1.134	1.216	1.285	1.343	1.393	1.474	1.538	1.590	1.655	1.737
5	1.284	1.444	1.581	1.699	1.802	1.973	2.113	2.231	2.379	2.574
4	1.440	1.674	1.878	2.055	2.211	2.476	2.695	2.882	3.120	3.438
3	1.772	2.154	2.492	2.792	3.060	3.521	3.910	4.246	4.680	5.275

Table 4.4.3 gives E_n and C_n for the distributions that Tukey (1960a) and Van Dyke (1961) have suggested may be used to approximate the normal and t-distributions. Van Dyke obtained the values of λ used in the approximations to the t-distributions by choosing λ so that the ratio of the 5 and .5 percent points is correct. The column in Table 4.4.3 headed "d.f.," meaning degrees of freedom, is used to identify the t-distribution being approximated.

The true values of E_n and C_n for the t-distribution with d.f. = 5,8, and 13, and $n=5$, have been given by Moses (1956) and are outlined below :

	E_n			C_n		
d.f.	5	8	13	5	8	13
True	.966	.986	.993	1.702	1.344	1.181
TRV	.970	.987	.994	1.699	1.343	1.186

Two approximations to the normal distribution (d.f. = ∞) are indicated in Table 4.4.3. Tukey (1960a) obtained the approximation

$$4.91\left(P^{.14} - (1-P)^{.14}\right)$$

to the 100 P'th percent points of the normal distribution by "cut and try methods," and the approximation

$$5.05\left(P^{.135} - (1-P)^{.135}\right)$$

is such that correct values are given for the 5 and .5 percent points. Some comparisons between these two approximations are given in Tables 4.4.3 (a) and (b), and in Table 4.4.4.

Table 4.4.4

Percent error in approximation to the P'th percent point of the normal distribution.

λ \ P	60	70	80	90	95	99	99.5	99.9
.14	-.4	-.4	-.2	-.0	.1	-.0	-.3	-1.5
.135	-.9	-.8	-.6	-.3	.0	.2	.0	.9

The agreement between the true and TRV-approximate values for E_5 and C_5 for the t-distributions is sufficient to indicate that the approximations may have wider applicability. Computations by Van Dyke indicate that the approximation to the percent points of the t-distributions is fairly good provided the degrees of freedom is not too small.

There seems to be little difference between the $\lambda = .14$ and $\lambda = .135$ approximations to the normal distribution. To the accuracy indicated in the tables, $\lambda = .135$ gives the correct value of β_2 and gives slightly more accurate values for C_n; $\lambda = .14$ gives better values for E_n. The percent points given by $\lambda = .14$ are more accurate for P between 60 and about 99 but the values for $\lambda = .135$ are more accurate for P=99.5 and 99.9.

Table 4.4.5

TRV distributions for which Plackett's upper bound on E_n is attained (λ = n-1).

(a) $E(R_n/d_n\sigma)$

λ+1	β_2	n = 2,3	4	5	6	8	10	12	15	20
2,3	1.80	1.023	1.010	.993	.976	.946	.921	.900	.873	.839
4	2.06	1.021	1.012	1.002	.992	.972	.955	.940	.920	.893
5	2.45	1.010	1.008	1.006	1.004	.998	.990	.983	.971	.953
6	2.90	.992	.996	1.002	1.007	1.015	1.018	1.018	1.015	1.007
8	3.88	.944	.960	.978	.996	1.026	1.047	1.062	1.076	1.086
10	4.88	.894	.917	.945	.972	1.019	1.054	1.081	1.110	1.138
12	5.88	.848	.876	.911	.944	1.003	1.049	1.086	1.127	1.172
15	7.38	.787	.821	.861	.901	.973	1.031	1.079	1.135	1.199
20	9.88	.708	.746	.791	.836	.919	.990	1.049	1.122	1.213

(b) $CV^2(R_n)$ / [$CV^2(R_n)$ from normal distribution]

λ+1	n = 2	3	4	5	6	8	10	12	15	20
2,3	.876	.726	.608	.518	.447	.345	.276	.228	.177	.126
4	.890	.787	.705	.636	.578	.484	.413	.357	.293	.222
5	.946	.908	.868	.826	.781	.693	.614	.545	.461	.358
6	1.046	1.077	1.081	1.062	1.029	.942	.851	.767	.659	.521
8	1.336	1.512	1.598	1.618	1.600	1.508	1.391	1.273	1.114	.902
10	1.692	2.019	2.179	2.232	2.226	2.125	1.981	1.832	1.623	1.335
12	2.078	2.560	2.793	2.878	2.881	2.772	2.604	2.424	2.168	1.808
15	2.686	3.406	3.747	3.879	3.899	3.780	3.578	3.356	3.033	2.568
20	3.736	4.858	5.383	5.597	5.649	5.522	5.271	4.985	4.559	3.928

Table 4.4.5 gives E_n and C_n for the values of λ for which Plackett's (1947) upper bound on $E(R_n / \sigma)$ is attained. There is no apparent manner in which these distributions are otherwise unusual or extraordinary.

All of the results given in this section are believed to be accurate to as many places as given.

Discussion of the "robustness" of E(R) as a Scale Parameter

The principal summary and analysis of the results on the range is deferred to Section 4.6 where the results of cox (1954), Burr (1967) and others (summarized in Section 4.3) are compared with those obtained here and in the following section.

It is convenient here to discuss the concept of the "robustness" of $E(R)$ as a scale parameter for symmetric Tukey random variables in the sense introduced in Section 4.1 above. More specifically consider the following. For a normal distribution

$$\mu \pm 1.96\ E(R_n / d_n)$$

covers the central 95 percent of the distribution. For any other symmetric distribution, define K_n to be such that

$$\mu \pm K_n\ E(R_n / d_n)$$

covers the central 95 percent of the distribution. Then the ratio $\rho_n = 1.96 / K_n$ may be considered a measure of the robustness of

$E(R_n)$ as a scale parameter for the central 95 percent of the distribution. More precisely the amount by which this ration differs from 1.0 is an index of the non-robustness of $E(R_n)$ as a scale parameter for that distribution.

A similar quantity may be defined for σ as $\rho = 1.96 / K$ where K is such that $\mu \pm K\sigma$ covers the central 95 percent of the distribution. Then ρ and ρ_n may be compared over a class of distributions for varying values of n and whichever deviates least from 1.0 (ratio-wise) could be said to be the more robust scale parameter for the central 95 percent of the distribution.

For Tukey random variables K_n may be computed quite simply as

$$K_n = z(.975) / E(R_n / d_n) ,$$

where $z(u)$ is defined in (4.4.1).

Table 4.4.6 gives ρ_n computed for the same values of n and β_2 (i.e. λ) as in Table 4.4.1. The value of ρ for these same values of β_2 is also given. The values of ρ and ρ_n give an indication of the robustness of $E(R_n)$ with respect to the central 95 percent of the distribution for symmetric Tukey random variables. These results may be compared with those given in Table 4.4.1 which may be interpreted as giving an indication of the robustness of $E(R_n / d_n)$ with respect to σ.

Table 4.4.6

"Robustness" of $E(R_n)$ as a scale parameter for the central 95 percent of the distribution. For details see text. Upper entry corresponds to the smaller value of λ.

β_2 =	2	3	4.2	6	8	10	15
σ	1.1164	.9958			.9679	.9716	.9795
	1.1187	.9350	.8335	.7602	.7266	.7180	.7535
n=2,3	1.1406	.9954			.9093	.9024	.8937
	1.1429	.9232	.7733	.6402	.5561	.5060	.4508
4	1.1297	.9956			.9184	.9124	.9050
	1.1323	.9288	.7886	.6623	.5814	.5330	.4803
5	1.1165	.9959			.9293	.9245	.9188
	1.1193	.9357	.8070	.6890	.6121	.5657	.5159
6	1.1034	.9961			.9405	.9368	.9327
	1.1063	.9419	.8248	.7151	.6425	.5983	.5521
8	1.0794	.9965			.9613	.9600	.9590
	1.0822	.9508	.8543	.7607	.6969	.6580	.6201
10	1.0588	.9968			.9799	.9806	.9825
	1.0610	.9555	.8761	.7969	.7422	.7089	.6805
12	1.0412	.9971			.9965	.9991	1.0035
	1.0425	.9572	.8919	.8255	.7794	.7521	.7339
15	1.0190	.9973			1.0182	1.0234	1.0314
	1.0187	.9563	.9077	.8578	.8237	.8050	.8026
20	.9900	.9975			1.0485	1.0573	1.0704
	.9870	.9501	.9215	.8930	.8761	.8706	.8938

From Table 4.4.6 it appears that for small values of n (less than about 8), 1.96σ would be a more robust scale parameter than $1.96\ E(R_n / d_n)$ whereas for larger values of n (about 8 to 15), the situation is reversed. All such conclusions are, of course, predicated on the class of distributions under consideration. If the only allowable shapes are restricted to those corresponding to β_2 equal 2 or 3 then $E(R_n / d_n)$ is better than σ for n larger than about 5.

In Table 4.4.1 it is interesting to note that for the values of λ included, E_n varies least for sample sizes of about 7 or 8. If only values of β_2 in the interval (3,9) are allowed then variation is least for $n = 9$ or 10. When the only allowable shapes correspond to small λ and β_2 in the interval (1.75, 4.20), E_n varies least for $n = 5$.

We make the empirical observation that E_n is a decreasing function of n for $\beta_2 = 2$ and is an increasing function of n for $\beta_2 \geq 3$. It may be noted that the relation

$$d_n^* = \frac{1}{\sqrt{\pi}} (1 + \lambda)(2 + \lambda) \frac{n}{n + \lambda} \left[\frac{1}{\lambda} - B(n,\lambda) \right]$$

where

$$\lambda = \frac{3(8 - 7\rho) + \sqrt{9\rho^2 + 96\rho - 128}}{2(3\rho - 4)}$$

and

$$\rho = \frac{2}{\pi} \arccos\left(- \frac{1}{3}\right)$$

gives d_n exactly for $n = 2(1)5$; is accurate to within 0.02% for $n = 6(1)10$; and is accurate to within 0.1% for $n = 11(1)20$. This value of $\lambda (=0.1416135)$ was found by identifying the relations for E_3 and E_4.

In Table 4.4.1 E_n differs substantially from 1.0 only for large values of λ and small values of n. Plackett's upper bound on E_n seems to increase indefinitely with n and for $n = 1000$ the upper bound on E_n is approximately 6.

These results on R are compared with those of Cox (1954) and others in Section 4.6.

4.5 PROPERTIES OF S, M, AND R FOR SOME DISCRETE DISTRIBUTIONS

We now direct attention to a type of non-normal distribution for which properties of S and M, as well as R, may readily be obtained for small samples. It has been indicated in Section 4.3 that there is virtually no information available on the properties of S and M in small samples from non-normal distributions. In this section we select a number of discrete distributions with from 3 to 11 equally-spaced possible values and compute the expectation and coefficient of variation for S, M, and R for samples of size 3, 5, and 10. These results are summarized at the end of the section and it is shown that the modified degrees of freedom approach due to Le Roux (1931) provides a good approximation to the expectation and coefficient of variation of S for the distributions considered here. For samples of size 3 and 5 this approach also gives a reasonably good approximation to the expectation of M and R; and for samples of size 3, 5, and 10 gives a good approximation to the coefficient of variation of M and R. The results on R obtained here are compared with those from other sources in Section 4.6.

Introduction

It is widely appreciated that real data are never truly continuous, but are in reality always discrete. Frequently the actual readings are so coarse that only 5 to 10 distinct values are ever obtained. Studies by Eisenhart (1947) and Taguti (1951) have investigated the effect of rounding on the distribution of S^2 . The results of such investigations are ordinarily used to justify the using of "continuous" techniques on "discrete" data. However, it would seem that they could also be used to partially justify the extrapolation of results on discrete populations to continuous populations. Even if this were not possible the properties obtained based on studies of discrete populations should certainly be applicable to real data. It would seem that the electronic computer could effectively be used to examine the properties of real-life data directly rather than using the digital computer to approximate a continuous function which in turn was only an approximation to discrete real-life data.

Discrete distributions have been used by a number of people to investigate the effect of departures from normality on the sampling distribution of various statistics. Rider (1929) investigated the behavior of $\bar{X}$, $\bar{X}/S$, the sample median, and R in samples from various discrete distributions with 5 and 10 categories for samples of size 2, 3, 4. Shone (1949) computed by hand the mean and variance of the sample range in samples from

18 discrete populations. Eleven of these had 7 categories; one had eight; two had nine; and four had ten. Burr (1955) gave a simplified method for computing the exact distribution of the range in samples from discrete populations and worked out two examples : an eleven category analogue of the normal distribution, and a twelve category analogue of a Pearson Type III distribution.

Tukey (1948) has even suggested that clues to the effect of non-normality "can probably be found by studying that extremely non-normal distribution in which all the probability is concentrated at two points." For such a random variable with $k=2$ distinct values it is easy to obtain rather simple analytic expressions for the expectation and coefficient of variation of S, M, and R. For larger values of k the analytic evaluation of these properties becomes much more difficult and it is more convenient to use the computer to systematically construct all possible samples, compute the designated statistics and probability for each sample and sum over all possible samples. Results on the expectation and coefficient of variation of S, M, and R were obtained for samples of size $n = 3,5,$ and 10 .

The enumeration of all of the possible k^n samples is feasible for small values of n and k but to extend the present investigation to larger values of n and k, methods for the reduction of the number of computations would have to be investigated. For symmetric distributions, for example, one could halve the computing time by taking advantage of symmetry.

In total 114 distributions were chosen for study although it was not economically feasible to do the computations for $n=10$ for the larger values of k. Thus for $n=10$ results were obtained for only 45 distributions all with $k \leq 7$ except for one with $k=8$. The distributions actually used are listed in Table 4.5.1 and some parent properties of the distributions are listed in Table 4.5.2. Table 4.5.3 lists the results of the computations on the expectation and coefficient of variation of S, M, and R.

The distributions numbered 37, 40, 42, 43, 45, 53, 60, 64, 66, 67, 88, 91, 92, 99, 100, 101, 103 and 106 are the ones for which Shone (1949) computed the expectation and standard deviation of the sample range for $n = 3, 4,$ and 5. (There appear to be a number of minor errors in his results, calculated on board ship by hand.) Distribution number 74 is Burr's (1955) discrete analogue of the normal distribution and number 108 is a modified version of his discrete analogue of a gamma distribution.

Table 4.5.1

Discrete Parent distributions: Relative frequencies

for each of the k equally spaced categories.

(Symmetric distributions)

ID	k	β_2	relative frequencies						
1	3	3.13	16	68	16				
2	3	3.50	1	5	1				
3	3	4.00	1	6	1				
4	3	4.50	1	7	1				
5	3	5.00	1	8	1				
6	3	5.50	1	9	1				
7	3	6.00	1	10	1				
8	4	2.44	11	39	39	11			
9	4	2.67	1	15	15	1			
10	4	2.71	7	43	43	7			
11	4	2.75	1	7	7	1			
12	4	2.76	1	11	11	1			
13	4	2.77	1	8	8	1			
14	4	2.77	1	10	10	1			
15	4	2.78	1	9	9	1			
16	5	1.70	1	1	1	1	1		
17	5	2.53	67	242	382	242	67		
18	5	2.79	1	10	20	10	1		
19	5	2.85	1	5	10	5	1		
20	5	2.94	12	215	546	215	12		
21	5	3.52	1	15	50	15	1		
22	5	4.15	1	5	20	5	1		
23	5	4.59	1	15	75	15	1		
24	5	4.78	1	10	50	10	1		
25	5	6.74	1	5	40	5	1		
26	5	8.04	1	5	50	5	1		
27	6	1.73	1	1	1	1	1	1	
28	6	2.59	1	1	3	3	1	1	
29	6	2.74	2	14	34	34	14	2	
30	6	2.86	1	10	30	30	10	1	
31	6	3.33	1	10	50	50	10	1	
32	6	3.37	1	5	20	20	5	1	
33	6	3.67	1	10	90	90	10	1	
34	6	3.82	1	5	30	30	5	1	
35	6	4.20	1	5	50	50	5	1	
36	7	1.75	143	143	143	142	143	143	143
37	7	1.78	9	9	9	10	9	9	9
38	7	2.17	100	120	150	260	150	120	100
39	7	2.32	60	130	180	260	180	130	60
40	7	2.49	2	7	15	16	15	7	2

Table 4.5.1 (Continued)

(Symmetric distributions, for k $\geq$ 8 only the first 7 frequencies are given)

ID	k	β_2	relative frequencies						
41	7	2.55	4	11	21	28	21	11	4
42	7	2.57	1	6	16	18	16	6	1
43	7	2.67	1	6	15	20	15	6	1
44	7	2.88	1	10	240	498	240	10	1
45	7	2.91	1	5	15	22	15	5	1
46	7	3.01	4	50	240	412	240	50	4
47	7	3.08	3	40	240	434	240	40	3
48	7	3.19	1	10	220	538	220	10	1
49	7	3.21	3	30	240	454	240	30	3
50	7	3.33	30	70	200	400	200	70	30
51	7	3.33	3	20	240	474	240	20	3
52	7	3.44	40	60	200	400	200	60	40
53	7	3.50	1	4	13	28	13	4	1
54	7	3.56	1	10	200	578	200	10	1
55	7	3.78	30	70	150	500	150	70	30
56	7	3.87	40	60	150	500	150	60	40
57	7	3.94	2	10	199	578	199	10	2
58	7	4.28	3	10	198	578	198	10	3
59	7	4.57	4	10	197	578	197	10	4
60	7	4.63	1	3	10	36	10	3	1
61	7	4.65	4	10	194	584	194	10	4
62	7	4.90	5	10	193	584	193	10	5
63	7	5.18	6	10	190	588	190	10	6
64	7	6.67	1	2	7	44	7	2	1
65	8	2.58	4	8	15	23	23	15	8
66	9	2.71	1	4	12	21	24	21	12
67	10	1.78	1	1	1	1	1	1	1
68	10	2.87	1	8	15	30	50	50	30
69	10	3.50	1	3	8	20	40	40	20
70	10	4.09	1	2	5	10	30	30	10
71	11	2.03	5	6	7	8	9	10	9
72	11	2.89	1	4	9	16	25	36	25
73	11	2.99	1	24	224	1109	2897	3989	2897
74	11	3.02	5	15	50	115	195	240	195
75	11	3.08	6	14	55	110	200	250	200
76	11	3.26	1	2	9	18	33	50	33
77	11	3.31	1	8	27	64	125	216	125
78	11	3.39	1	64	729	2744	6859	13824	6859
79	11	6.21	2	20	100	500	800	6000	800
80	11	9.40	1	4	20	100	160	2000	160
81	11	10.07	3	6	8	10	190	588	190
82	11	13.28	3	6	8	10	120	700	120
83	11	16.87	1	2	3	10	100	800	100
84	11	20.13	1	2	3	10	100	1000	100
85	11	29.72	1	2	3	10	50	1000	50

Table 4.5.1 (Continued)

(Non-symmetric distributions)

ID	k	β_2	relative frequencies										
86	3	2.61	7	62	31								
87	6	2.52	7	17	30	28	14	4					
88	7	2.70	1	4	9	15	20	10	5				
89	7	3.17	30	70	70	150	150	500	30				
90	7	3.21	30	70	150	150	500	70	30				
91	7	3.22	1	2	4	8	12	16	21				
92	7	3.44	1	3	6	12	25	15	2				
93	7	3.59	10	20	80	200	350	300	40				
94	7	3.61	30	30	70	70	150	150	500				
95	7	4.20	10	50	100	300	500	30	10				
96	7	5.39	20	20	40	60	100	200	560				
97	7	11.71	1	1	1	1	10	20	67				
98	8	2.55	3	7	14	22	23	17	9	5			
99	8	3.05	46	234	308	233	120	45	12	2			
100	9	2.99	35	188	275	244	154	72	25	6	1		
101	10	2.76	1	10	44	117	205	246	205	117	44	10	
102	10	4.49	1	3	5	7	15	30	50	80	40	5	
103	10	4.54	134	307	250	149	81	43	19	10	5	2	
104	10	4.99	1	0	3	10	30	70	30	10	3	1	
105	10	5.41	100	60	37	22	13	8	5	3	2	1	
106	10	6.98	1	2	4	8	17	34	67	133	245	489	
107	10	8.88	1	1	2	3	5	10	50	100	70	5	
108	11	3.59	1	13	22	21	17	11	7	4	2	1	1
109	11	6.17	100	60	37	22	13	8	5	3	2	1	1
110	11	6.47	150	90	55	32	18	10	7	4	3	2	1
111	11	7.44	1	2	3	4	10	50	300	500	1000	100	1
112	11	8.15	1	2	3	4	10	100	500	1000	300	12	1
113	11	8.54	1	1	1	3	5	10	50	100	70	10	1
114	11	19.45	1	2	3	1	50	500	50	10	3	2	1

Table 4.5.2

Some properties of the discrete parent distributions.

(ν_j denotes $E|X-\mu|^j$.)

ID	k	β_2	δ/σ	ν_3/σ^3	ν_5/σ^5	β_1	μ	σ
1	3	3.13	.5657	1.768	5.5	.000	2.000	.566
2	3	3.50	.5345	1.871	6.5	.000	2.000	.535
3	3	4.00	.5000	2.000	8.0	.000	2.000	.500
4	3	4.50	.4714	2.121	9.5	.000	2.000	.471
5	3	5.00	.4472	2.236	11.2	.000	2.000	.447
6	3	5.50	.4264	2.345	12.9	.000	2.000	.426
7	3	6.00	.4082	2.449	14.7	.000	2.000	.408
8	4	2.44	.8668	1.466	4.3	.000	2.500	.831
9	4	2.67	.9186	1.429	5.9	.000	2.500	.612
10	4	2.71	.8791	1.503	5.3	.000	2.500	.728
11	4	2.75	.8839	1.503	5.5	.000	2.500	.707
12	4	2.76	.9037	1.472	5.9	.000	2.500	.645
13	4	2.77	.8893	1.498	5.7	.000	2.500	.687
14	4	2.77	.8992	1.482	5.9	.000	2.500	.657
16	5	1.70	.8485	1.273	2.3	.000	3.000	1.414
17	5	2.53	.7446	1.510	4.5	.000	3.000	1.010
18	5	2.79	.6999	1.575	5.5	.000	3.000	.816
19	5	2.85	.7035	1.597	5.6	.000	3.000	.905
20	5	2.94	.6591	1.630	6.0	.000	3.000	.725
21	5	3.52	.6091	1.778	7.8	.000	3.000	.681
22	5	4.15	.5833	1.926	9.7	.000	3.000	.750
23	5	4.59	.5332	2.031	11.7	.000	3.000	.596
24	5	4.78	.5345	2.062	12.4	.000	3.000	.624
25	5	6.74	.4576	2.455	20.2	.000	3.000	.588
26	5	8.04	.4191	2.681	26.3	.000	3.000	.539
27	6	1.73	.8783	1.280	2.4	.000	3.500	1.708
28	6	2.59	.8087	1.540	4.5	.000	3.500	1.360
29	6	2.74	.8393	1.538	5.4	.000	3.500	1.025
30	6	2.86	.8430	1.559	5.8	.000	3.500	.940
31	6	3.33	.8473	1.640	7.6	.000	3.500	.822
32	6	3.37	.8269	1.672	7.5	.000	3.500	.930
33	6	3.67	.8687	1.661	9.5	.000	3.500	.712
34	6	3.82	.8333	1.740	9.4	.000	3.500	.833
35	6	4.20	.8539	1.765	11.7	.000	3.500	.732
36	7	1.75	.8576	1.285	2.5	.000	4.000	2.001
37	7	1.78	.8504	1.296	2.5	.000	4.000	1.984
38	7	2.17	.7889	1.424	3.5	.000	4.000	1.749
39	7	2.32	.7874	1.454	3.9	.000	4.000	1.575
40	7	2.49	.7922	1.484	4.5	.000	4.000	1.381

Table 4.5.2 (Continued)

ID	k	β_2	δ/σ	ν_3/σ^3	ν_5/σ^5	β_1	μ	σ
41	7	2.55	.7740	1.512	4.6	.000	4.000	1.421
42	7	2.57	.7829	1.501	4.9	.000	4.000	1.237
43	7	2.67	.7655	1.531	5.1	.000	4.000	1.225
44	7	2.88	.6919	1.579	6.3	.000	4.000	.760
45	7	2.91	.7462	1.589	5.9	.000	4.000	1.173
46	7	3.01	.7215	1.611	6.4	.000	4.000	.976
47	7	3.08	.7120	1.624	6.7	.000	4.000	.924
48	7	3.19	.6626	1.657	7.4	.000	4.000	.733
49	7	3.21	.7025	1.648	7.3	.000	4.000	.880
50	7	3.33	.7022	1.709	7.1	.000	4.000	1.225
51	7	3.33	.6938	1.664	8.0	.000	4.000	.833
52	7	3.44	.6957	1.739	7.3	.000	4.000	1.265
53	7	3.50	.6882	1.739	7.8	.000	4.000	1.090
54	7	3.56	.6320	1.747	8.7	.000	4.000	.706
55	7	3.78	.6423	1.835	8.3	.000	4.000	1.183
56	7	3.87	.6369	1.862	8.6	.000	4.000	1.225
57	7	3.94	.6277	1.807	10.6	.000	4.000	.717
58	7	4.28	.6236	1.861	12.2	.000	4.000	.728
59	7	4.57	.6198	1.909	13.5	.000	4.000	.739
60	7	4.63	.6033	1.999	11.8	.000	4.000	.984
61	7	4.65	.6151	1.925	13.9	.000	4.000	.735
62	7	4.90	.6115	1.968	15.0	.000	4.000	.746
63	7	5.18	.6050	2.018	16.2	.000	4.000	.754
64	7	6.67	.5052	2.406	20.1	.000	4.000	.866
65	8	2.58	.8171	1.518	4.7	.000	4.500	1.664
66	9	2.71	.7842	1.535	5.3	.000	5.000	1.556
67	10	1.78	.8704	1.292	2.5	.000	5.500	2.872
68	10	2.87	.8034	1.586	5.6	.000	5.500	1.676
69	10	3.50	.7870	1.717	8.0	.000	5.500	1.500
70	10	4.09	.7570	1.857	9.9	.000	5.500	1.486
71	11	2.03	.8332	1.372	3.2	.000	6.000	2.850
72	11	2.89	.7636	1.593	5.7	.000	6.000	1.884
73	11	2.99	.7540	1.597	6.3	.000	6.000	1.250
74	11	3.02	.7698	1.607	6.3	.000	6.000	1.715
75	11	3.08	.7649	1.622	6.5	.000	6.000	1.720
76	11	3.26	.7454	1.667	7.1	.000	6.000	1.662
77	11	3.31	.7294	1.688	7.2	.000	6.000	1.528
78	11	3.39	.7050	1.709	7.5	.000	6.000	1.212
79	11	6.21	.5162	2.313	18.9	.000	6.000	.959
80	11	9.40	.4273	2.825	35.9	.000	6.000	.803
81	11	10.07	.5553	2.736	42.7	.000	6.000	.962
82	11	13.28	.4536	3.238	60.8	.000	6.000	.900
83	11	16.87	.4177	3.483	98.7	.000	6.000	.659
84	11	20.13	.3823	3.805	128.7	.000	6.000	.603
85	11	29.72	.2932	4.810	211.5	.000	6.000	.554

Table 4.5.2 (Continued)

ID	k	β_2	δ/σ	ν_3/σ^3	ν_5/σ^5	β_1	μ	σ
86	3	2.61	.8299	1.519	4.8	.001	2.240	.568
87	6	2.52	.8289	1.498	4.5	.000	3.370	1.230
88	7	2.70	.8189	1.541	5.1	.060	4.547	1.369
89	7	3.17	.8183	1.613	7.0	1.147	4.940	1.468
90	7	3.21	.8152	1.648	6.9	.456	4.350	1.276
91	7	3.22	.8287	1.611	7.5	.850	5.500	1.490
92	7	3.44	.7824	1.697	7.9	.579	4.719	1.231
93	7	3.59	.7594	1.701	9.0	.519	4.920	1.146
94	7	3.61	.8196	1.698	8.7	1.535	5.730	1.648
95	7	4.20	.8032	1.830	10.9	.799	4.360	.985
96	7	5.39	.7458	2.043	16.0	2.931	6.040	1.442
97	7	11.71	.7213	2.891	52.8	7.169	6.426	1.056
98	8	2.55	.8163	1.510	4.6	.001	4.670	1.662
99	8	3.05	.8158	1.595	6.7	.258	3.340	1.289
100	9	2.99	.8189	1.580	6.5	.231	3.646	1.416
101	10	2.76	.7798	1.549	5.4	.001	5.995	1.577
102	10	4.49	.7700	1.874	12.5	1.290	7.225	1.617
103	10	4.54	.7528	1.856	13.4	1.270	3.035	1.592
104	10	4.99	.6614	1.992	14.8	.039	5.994	1.220
105	10	5.41	.7742	2.012	17.0	2.469	2.466	1.797
106	10	6.98	.7318	2.211	26.8	3.317	8.966	1.405
107	10	8.88	.6880	2.503	37.3	3.516	7.838	1.275
108	11	3.59	.8008	1.697	8.9	.737	4.480	1.957
109	11	6.17	.7632	2.124	21.1	2.911	2.500	1.872
110	11	6.47	.7566	2.175	22.4	3.131	2.452	1.836
111	11	7.44	.8062	2.122	36.1	2.006	8.367	1.006
112	11	8.15	.7453	2.240	41.6	1.241	7.766	.881
113	11	8.54	.6769	2.442	36.1	2.680	7.913	1.276
114	11	19.45	.4090	3.874	111.2	.042	6.029	.707

Table 4.5.3a

The standardized expectation and coefficient of variation of S, M and R for samples from the discrete distributions.

n = 3

ID	k	β_2	n^*	E(S)	E(M)	E(M/δ)	E(R)	C(S)	C(M)	C(R)
1	3	3.13	2.92	.896	.909	1.282	.842	2.145	1.997	2.312
2	3	3.50	2.71	.866	.881	1.315	.811	2.552	2.388	2.723
3	3	4.00	2.50	.831	.847	1.352	.775	3.096	2.931	3.284
4	3	4.50	2.33	.798	.817	1.382	.743	3.652	3.490	3.857
5	3	5.00	2.20	.768	.788	1.405	.714	4.235	4.059	4.437
6	3	5.50	2.09	.742	.762	1.425	.688	4.811	4.636	5.022
7	3	6.00	2.00	.717	.737	1.441	.664	5.405	5.218	5.610
8	4	2.44	3.46	.991	.992	.913	.951	1.089	1.034	1.208
9	4	2.67	3.25	.966	.990	.860	.899	1.332	1.249	1.466
10	4	2.71	3.21	.975	.984	.893	.926	1.238	1.156	1.381
11	4	2.75	3.18	.973	.984	.888	.921	1.263	1.179	1.409
12	4	2.76	3.17	.968	.986	.871	.906	1.317	1.232	1.464
13	4	2.77	3.17	.971	.984	.883	.916	1.285	1.198	1.433
14	4	2.77	3.17	.968	.985	.874	.909	1.314	1.223	1.458
15	4	2.78	3.16	.969	.984	.878	.911	1.307	1.213	1.448
16	5	1.70	4.53	1.027	1.031	.969	1.003	.755	.771	.787
17	5	2.53	3.37	.998	.996	1.067	.969	1.022	.999	1.123
18	5	2.79	3.15	.981	.983	1.120	.941	1.184	1.130	1.322
19	5	2.85	3.11	.981	.981	1.112	.947	1.185	1.135	1.305
20	5	2.94	3.04	.962	.967	1.170	.916	1.380	1.296	1.532
21	5	3.52	2.70	.927	.936	1.226	.880	1.758	1.649	1.923
22	5	4.15	2.45	.906	.913	1.249	.863	2.023	1.931	2.198
23	5	4.59	2.31	.862	.875	1.310	.811	2.612	2.479	2.803
24	5	4.78	2.26	.862	.875	1.306	.813	2.613	2.480	2.797
25	5	6.74	1.89	.777	.794	1.384	.729	4.050	3.921	4.263
26	5	8.04	1.75	.728	.746	1.420	.679	5.136	4.978	5.337
27	6	1.73	4.47	1.028	1.032	.937	1.009	.749	.767	.767
28	6	2.59	3.32	.997	.996	.983	.977	1.028	1.014	1.107
29	6	2.74	3.19	.993	.992	.943	.965	1.063	1.037	1.163
30	6	2.86	3.10	.988	.987	.935	.955	1.117	1.083	1.237
31	6	3.33	2.80	.970	.975	.918	.928	1.296	1.234	1.445
32	6	3.37	2.78	.972	.974	.940	.937	1.274	1.228	1.412
33	6	3.67	2.63	.957	.971	.892	.903	1.428	1.344	1.592
34	6	3.82	2.57	.958	.965	.924	.915	1.422	1.355	1.582

Table 4.5.3a (Continued)

ID	k	β_2	n^*	E(S)	E(M)	E(M/δ)	E(R)	C(S)	C(M)	C(R)
35	6	4.20	2.43	.947	.962	.899	.894	1.540	1.450	1.707
36	7	1.75	4.43	1.029	1.032	.960	1.013	.744	.762	.756
37	7	1.78	4.37	1.028	1.031	.968	1.013	.747	.766	.763
38	7	2.17	3.76	1.016	1.015	1.026	1.003	.855	.859	.896
39	7	2.32	3.59	1.012	1.010	1.023	.998	.891	.893	.938
40	7	2.49	3.41	1.007	1.005	1.012	.990	.935	.933	.996
41	7	2.55	3.35	1.005	1.002	1.033	.989	.956	.957	1.022
42	7	2.57	3.33	1.004	1.002	1.022	.983	.960	.959	1.038
43	7	2.67	3.25	1.000	.998	1.040	.979	.997	.988	1.075
44	7	2.88	3.08	.977	.981	1.131	.932	1.220	1.151	1.357
45	7	2.91	3.06	.992	.990	1.058	.969	1.074	1.057	1.162
46	7	3.01	2.99	.985	.984	1.088	.954	1.146	1.111	1.258
47	7	3.08	2.95	.981	.981	1.100	.948	1.178	1.139	1.302
48	7	3.19	2.88	.962	.967	1.165	.917	1.374	1.296	1.527
49	7	3.21	2.87	.977	.978	1.110	.940	1.227	1.176	1.353
50	7	3.33	2.80	.974	.973	1.106	.952	1.250	1.226	1.348
51	7	3.33	2.80	.973	.975	1.122	.933	1.263	1.206	1.398
52	7	3.44	2.74	.970	.969	1.112	.948	1.295	1.270	1.395
53	7	3.50	2.71	.968	.968	1.122	.942	1.314	1.281	1.428
54	7	3.56	2.69	.942	.949	1.198	.896	1.590	1.491	1.745
55	7	3.78	2.59	.945	.948	1.178	.919	1.557	1.519	1.678
56	7	3.87	2.55	.940	.944	1.182	.914	1.611	1.568	1.728
57	7	3.94	2.52	.938	.944	1.201	.892	1.642	1.549	1.802
58	7	4.28	2.40	.933	.940	1.203	.888	1.693	1.603	1.856
59	7	4.57	2.31	.929	.936	1.205	.884	1.743	1.654	1.906
60	7	4.63	2.30	.915	.921	1.218	.882	1.901	1.848	2.053
61	7	4.65	2.29	.925	.933	1.210	.880	1.783	1.692	1.947
62	7	4.90	2.22	.921	.929	1.212	.877	1.833	1.741	1.995
63	7	5.18	2.16	.915	.923	1.217	.871	1.905	1.815	2.069
64	7	6.67	1.90	.826	.839	1.326	.786	3.166	3.083	3.343
65	8	2.58	3.33	1.005	1.003	.979	.993	.953	.957	1.006
66	9	2.71	3.21	1.003	1.001	1.018	.989	.972	.972	1.026
67	10	1.78	4.37	1.029	1.033	.947	1.018	.740	.759	.741
68	10	2.87	3.09	.997	.994	.987	.985	1.033	1.029	1.086
69	10	3.50	2.71	.980	.978	.992	.965	1.191	1.186	1.265
70	10	4.09	2.47	.959	.959	1.011	.941	1.409	1.396	1.497
71	11	2.03	3.96	1.023	1.024	.981	1.016	.792	.805	.802
72	11	2.89	3.08	.997	.995	1.039	.987	1.029	1.027	1.074
73	11	2.99	3.01	.993	.991	1.049	.973	1.064	1.053	1.143
74	11	3.02	2.99	.996	.993	1.030	.984	1.040	1.040	1.091

Table 4.5.3a (Continued)

n = 3

ID	k	β_2	n^*	E(S)	E(n)	E(M/δ)	E(R)	C(S)	C(M)	C(R)
75	11	3.08	2.95	.994	.992	1.034	.982	1.056	1.058	1.110
76	11	3.26	2.84	.988	.986	1.055	.975	1.116	1.113	1.174
77	11	3.31	2.81	.983	.981	1.073	.968	1.163	1.153	1.231
78	11	3.39	2.77	.976	.975	1.103	.954	1.230	1.214	1.333
79	11	6.21	1.97	.840	.852	1.317	.801	2.942	2.868	3.110
80	11	9.40	1.64	.736	.754	1.408	.692	4.931	4.827	5.114
81	11	10.07	1.60	.853	.863	1.240	.814	2.747	2.720	2.901
82	11	13.28	1.45	.756	.772	1.357	.714	4.504	4.461	4.679
83	11	16.87	1.36	.716	.734	1.402	.671	5.417	5.320	5.602
84	11	20.13	1.30	.670	.689	1.437	.624	6.719	6.611	6.907
85	11	29.72	1.34	.621	.584	1.590	.577	14.715	13.896	14.855
86	3	2.61	3.30	.935	.957	.920	.871	1.668	1.556	1.800
87	6	2.52	3.38	1.003	1.001	.963	.983	.969	.963	1.047
88	7	2.70	3.22	.999	.996	.971	.981	1.014	1.007	1.080
89	7	3.17	2.89	.948	.959	.936	.913	1.528	1.540	1.568
90	7	3.21	2.87	.956	.963	.942	.925	1.443	1.425	1.522
91	7	3.22	2.86	.979	.984	.947	.953	1.207	1.217	1.247
92	7	3.44	2.74	.970	.972	.991	.946	1.289	1.282	1.369
93	7	3.59	2.67	.976	.977	1.027	.949	1.231	1.225	1.322
94	7	3.61	2.66	.935	.948	.923	.899	1.671	1.684	1.701
95	7	4.20	2.43	.934	.944	.938	.895	1.678	1.642	1.707
96	7	5.39	2.11	.877	.894	.956	.837	2.401	2.403	2.436
97	7	11.71	1.51	.784	.803	.888	.736	3.922	3.926	3.999
98	8	2.55	3.35	1.006	1.003	.981	.994	.946	.947	.995
99	8	3.05	2.97	.992	.991	.969	.971	1.076	1.079	1.151
100	9	2.99	3.01	.995	.994	.968	.977	1.046	1.050	1.106
101	10	2.76	3.17	1.002	.999	1.022	.988	.986	.985	1.039
102	10	4.49	2.34	.952	.955	.990	.932	1.480	1.503	1.525
103	10	4.54	2.32	.959	.961	1.019	.939	1.409	1.432	1.450
104	10	4.99	2.20	.942	.943	1.138	.918	1.592	1.570	1.696
105	10	5.41	2.11	.918	.929	.957	.890	1.869	1.895	1.892
106	10	6.98	1.86	.892	.905	.987	.856	2.191	2.205	2.235
107	10	8.88	1.68	.893	.899	1.043	.863	2.185	2.221	2.258
108	11	3.59	2.67	.980	.981	.978	.967	1.192	1.212	1.219
109	11	6.17	1.97	.907	.918	.960	.879	2.000	2.031	2.017
110	11	6.47	1.93	.901	.913	.962	.872	2.077	2.110	2.096
111	11	7.44	1.81	.924	.932	.923	.885	1.792	1.787	1.880
112	11	8.15	1.74	.923	.928	.993	.884	1.812	1.780	1.947
113	11	8.54	1.70	.904	.909	1.071	.877	2.036	2.062	2.111
114	11	19.45	1.31	.660	.679	1.325	.616	7.029	6.944	7.204

Table 4.5.3b

The standardized expectation and coefficient of variation of S, M and R for samples from the discrete distributions.

n = 5

ID	k	β_2	n^*	E(S)	E(M)	E(M/δ)	E(R)	C(S)	C(M)	C(R)
1	3	3.13	4.80	.950	.915	1.291	.884	1.930	1.949	2.291
2	3	3.50	4.33	.928	.889	1.327	.864	2.386	2.357	2.768
3	3	4.00	3.86	.901	.859	1.370	.838	2.999	2.935	3.422
4	3	4.50	3.50	.875	.831	1.406	.813	3.639	3.541	4.091
5	3	5.00	3.22	.849	.805	1.435	.788	4.326	4.168	4.771
6	3	5.50	3.00	.826	.781	1.462	.765	5.007	4.812	5.458
7	3	6.00	2.82	.803	.759	1.483	.743	5.719	5.464	6.151
8	4	2.44	6.15	1.006	1.014	.934	.942	.892	.936	1.152
9	4	2.67	5.61	1.003	1.046	.909	.865	.950	.914	1.452
10	4	2.71	5.52	1.001	1.017	.923	.913	.989	.976	1.409
11	4	2.75	5.44	1.000	1.020	.920	.906	.997	.974	1.448
12	4	2.76	5.42	1.001	1.034	.913	.880	.981	.946	1.495
13	4	2.77	5.41	1.000	1.023	.918	.898	1.000	.969	1.476
14	4	2.77	5.41	1.000	1.030	.914	.885	.997	.954	1.496
15	4	2.78	5.39	.999	1.026	.915	.891	1.009	.962	1.491
16	5	1.70	9.33	1.026	1.055	.992	.963	.565	.683	.531
17	5	2.53	5.93	1.006	1.007	1.079	.964	.892	.951	1.060
18	5	2.79	5.37	1.000	.993	1.132	.941	1.008	1.075	1.277
19	5	2.85	5.26	.997	.988	1.121	.953	1.057	1.094	1.286
20	5	2.94	5.10	.990	.974	1.179	.927	1.179	1.247	1.486
21	5	3.52	4.31	.967	.940	1.232	.911	1.593	1.617	1.951
22	5	4.15	3.74	.949	.917	1.255	.909	1.942	1.926	2.320
23	5	4.59	3.44	.921	.883	1.322	.867	2.530	2.487	2.964
24	5	4.78	3.34	.920	.882	1.317	.869	2.558	2.497	2.983
25	5	6.74	2.60	.852	.810	1.412	.805	4.240	4.078	4.708
26	5	8.04	2.33	.809	.767	1.460	.761	5.529	5.263	5.965
27	6	1.73	9.13	1.025	1.054	.958	.972	.578	.689	.527
28	6	2.59	5.78	1.003	.996	.983	.980	.947	1.004	1.048
29	6	2.74	5.46	1.003	1.003	.954	.962	.953	.982	1.141
30	6	2.86	5.24	1.001	1.001	.948	.953	.987	1.011	1.241
31	6	3.33	4.53	.992	.995	.937	.926	1.144	1.095	1.556
32	6	3.37	4.48	.990	.985	.951	.943	1.179	1.138	1.518
33	6	3.67	4.15	.988	1.006	.924	.891	1.211	1.090	1.774
34	6	3.82	4.01	.983	.983	.941	.918	1.307	1.194	1.785

Table 4.5.3b (Continued)

n = 5

ID	k	β_2	n^*	E(S)	E(M)	E(M/δ)	E(R)	C(S)	C(M)	C(R)
35	6	4.20	3.70	.979	.993	.928	.885	1.374	1.179	1.982
36	7	1.75	9.00	1.025	1.054	.980	.978	.582	.690	.524
37	7	1.78	8.81	1.025	1.051	.986	.979	.590	.699	.538
38	7	2.17	6.99	1.015	1.022	1.034	.990	.750	.835	.756
39	7	2.32	6.49	1.012	1.018	1.031	.989	.800	.867	.833
40	7	2.49	6.03	1.009	1.015	1.022	.981	.848	.894	.929
41	7	2.55	5.88	1.007	1.008	1.039	.985	.880	.933	.963
42	7	2.57	5.83	1.008	1.014	1.033	.974	.861	.910	.987
43	7	2.67	5.61	1.005	1.006	1.049	.975	.914	.951	1.036
44	7	2.88	5.20	.999	.991	1.143	.930	1.021	1.088	1.289
45	7	2.91	5.15	1.000	.996	1.065	.970	1.007	1.024	1.163
46	7	3.01	4.98	.997	.992	1.097	.956	1.046	1.067	1.266
47	7	3.08	4.88	.996	.990	1.109	.950	1.065	1.092	1.314
48	7	3.19	4.72	.989	.974	1.173	.926	1.185	1.247	1.498
49	7	3.21	4.69	.994	.986	1.120	.943	1.110	1.127	1.375
50	7	3.33	4.53	.987	.972	1.104	.969	1.227	1.224	1.404
51	7	3.33	4.53	.992	.984	1.132	.935	1.130	1.153	1.417
52	7	3.44	4.40	.984	.966	1.108	.968	1.287	1.277	1.468
53	7	3.50	4.33	.984	.968	1.122	.960	1.281	1.271	1.500
54	7	3.56	4.27	.976	.954	1.204	.917	1.427	1.453	1.760
55	7	3.78	4.05	.970	.944	1.172	.954	1.540	1.531	1.763
56	7	3.87	3.97	.966	.938	1.175	.952	1.613	1.590	1.826
57	7	3.94	3.91	.972	.949	1.206	.914	1.504	1.517	1.856
58	7	4.28	3.65	.968	.944	1.208	.912	1.580	1.578	1.945
59	7	4.57	3.46	.964	.939	1.209	.910	1.656	1.636	2.028
60	7	4.63	3.42	.950	.919	1.216	.926	1.920	1.869	2.235
61	7	4.65	3.41	.962	.936	1.214	.908	1.701	1.676	2.077
62	7	4.90	3.27	.958	.931	1.215	.905	1.776	1.731	2.154
63	7	5.18	3.14	.953	.925	1.220	.902	1.872	1.813	2.260
64	7	6.67	2.62	.886	.845	1.335	.855	3.348	3.194	3.730
65	8	2.58	5.81	1.006	1.007	.983	.991	.894	.942	.952
66	9	2.71	5.52	1.005	1.007	1.025	.985	.919	.945	.903
67	10	1.78	8.81	1.025	1.053	.965	.986	.589	.695	.521
68	10	2.87	5.22	1.000	.995	.988	.988	1.005	1.023	1.077
69	10	3.50	4.33	.988	.977	.990	.976	1.211	1.182	1.348
70	10	4.09	3.79	.973	.952	1.004	.963	1.481	1.411	1.664
71	11	2.03	7.54	1.019	1.035	.991	.996	.683	.770	.643
72	11	2.89	5.18	.999	.995	1.039	.991	1.010	1.026	1.065
73	11	2.99	5.02	.999	.997	1.055	.973	1.012	1.021	1.150
74	11	3.02	4.97	.999	.996	1.032	.986	1.025	1.027	1.101

Table 4.5.3b (Continued)

n = 5

ID	k	β_2	n^*	E(S)	E(M)	E(M/δ)	E(R)	C(S)	C(M)	C(R)
75	11	3.08	4.88	.997	.993	1.036	.986	1.045	1.048	1.128
76	11	3.26	4.62	.993	.985	1.054	.983	1.121	1.112	1.213
77	11	3.31	4.56	.990	.979	1.071	.980	1.167	1.156	1.277
78	11	3.39	4.46	.988	.975	1.103	.969	1.202	1.208	1.386
79	11	6.21	2.75	.898	.858	1.327	.870	3.063	2.942	3.413
80	11	9.40	2.12	.814	.771	1.441	.776	5.388	5.111	5.756
81	11	10.07	2.04	.895	.860	1.235	.858	3.139	2.847	3.555
82	11	13.28	1.78	.820	.776	1.366	.785	5.183	4.828	5.578
83	11	16.87	1.61	.791	.748	1.430	.747	6.141	5.757	6.551
84	11	20.13	1.51	.750	.708	1.477	.704	7.697	7.234	8.080
85	11	29.72	1.20	.538	.557	1.515	.497	12.453	12.331	12.631
86	3	2.61	5.74	.982	.994	.955	.867	1.317	1.318	1.730
87	6	2.52	5.95	1.008	1.009	.972	.978	.872	.928	.983
88	7	2.70	5.55	1.003	1.002	.976	.980	.950	.988	1.035
89	7	3.17	4.75	.976	.992	.968	.903	1.427	1.471	1.455
90	7	3.21	4.69	.980	.981	.960	.935	1.350	1.387	1.495
91	7	3.22	4.68	.990	1.002	.964	.937	1.172	1.173	1.220
92	7	3.44	4.40	.984	.978	.997	.953	1.274	1.293	1.393
93	7	3.59	4.24	.988	.986	1.036	.948	1.205	1.193	1.379
94	7	3.61	4.22	.966	.982	.956	.886	1.618	1.609	1.644
95	7	4.20	3.70	.965	.963	.957	.901	1.642	1.537	1.961
96	7	5.39	3.04	.922	.927	.992	.842	2.524	2.412	2.571
97	7	11.71	1.89	.845	.846	.936	.757	4.453	3.954	4.720
98	8	2.55	5.88	1.007	1.008	.986	.990	.886	.929	.936
99	8	3.05	4.92	.998	1.000	.978	.967	1.026	1.051	1.144
100	9	2.99	5.02	1.000	1.002	.976	.972	1.004	1.025	1.090
101	10	2.76	5.42	1.004	1.004	1.028	.985	.939	.962	1.012
102	10	4.49	3.51	.967	.960	.995	.943	1.593	1.554	1.644
103	10	4.54	3.48	.972	.968	1.026	.944	1.511	1.463	1.561
104	10	4.99	3.23	.962	.939	1.133	.945	1.688	1.591	1.908
105	10	5.41	3.04	.944	.947	.976	.889	2.042	1.931	2.084
106	10	6.98	2.54	.927	.931	1.015	.857	2.407	2.200	2.536
107	10	8.88	2.19	.919	.902	1.046	.882	2.583	2.338	2.776
108	11	3.59	4.24	.987	.988	.984	.965	1.233	1.221	1.252
109	11	6.17	2.76	.935	.936	.978	.881	2.236	2.090	2.276
110	11	6.47	2.68	.930	.930	.980	.875	2.341	2.176	2.392
111	11	7.44	2.44	.953	.959	.949	.886	1.863	1.710	2.113
112	11	8.15	2.31	.951	.936	1.002	.903	1.905	1.788	2.237
113	11	8.54	2.24	.928	.910	1.073	.897	2.388	2.151	2.588
114	11	19.45	1.53	.739	.697	1.360	.696	8.156	7.654	8.489

Table 4.5.3c
The standardized expectation and coefficient..., etc.
n = 10

ID	k	β_2	n^*	E(S)	E(M)	E(M/δ)	E(R)	C(S)	C(M)	C(R)
1	3	3.13	9.5	.989	.881	1.243	.948	1.405	2.261	1.477
8	4	2.44	13.0	1.006	1.038	.956	.929	.787	.850	1.071
10	4	2.71	11.3	1.003	1.049	.953	.906	.885	.823	1.724
17	5	2.53	12.4	1.005	1.005	1.077	.949	.823	.975	.955
20	5	2.94	10.2	.999	.953	1.154	.930	1.019	1.415	1.240
27	6	1.73	22.0	1.015	1.072	.974	.883	.459	.629	.240
28	6	2.59	12.0	1.003	.993	.980	.976	.882	1.032	.765
29	6	2.74	11.2	1.003	1.014	.964	.956	.904	.944	1.106
36	7	1.75	21.6	1.015	1.071	.997	.892	.466	.632	.249
37	7	1.78	21.0	1.014	1.067	1.001	.897	.480	.647	.260
38	7	2.17	15.3	1.009	1.024	1.036	.950	.670	.844	.472
39	7	2.32	14.0	1.007	1.021	1.035	.961	.730	.867	.635
40	7	2.49	12.7	1.006	1.023	1.030	.962	.794	.871	.855
41	7	2.55	12.3	1.004	1.010	1.042	.972	.835	.935	.837
42	7	2.57	12.1	1.005	1.020	1.040	.956	.824	.886	.969
43	7	2.67	11.6	1.003	1.009	1.052	.964	.872	.949	.989
44	7	2.88	10.5	1.002	.978	1.127	.905	.921	1.200	1.191
45	7	2.91	10.4	1.001	.996	1.065	.972	.980	1.038	1.164
46	7	3.01	10.0	1.000	.987	1.091	.955	1.011	1.113	1.316
47	7	3.08	9.7	.999	.982	1.101	.949	1.033	1.153	1.387
48	7	3.19	9.3	.998	.955	1.150	.924	1.066	1.406	1.345
49	7	3.21	9.2	.994	.992	.970	.959	1.220	1.429	1.361
50	7	3.33	8.8	.994	.961	1.092	1.001	1.212	1.295	1.326
51	7	3.33	8.8	.997	.973	1.118	.928	1.097	1.245	1.549
52	7	3.44	8.5	.993	.954	1.094	1.006	1.278	1.354	1.379
53	7	3.50	8.3	.993	.956	1.108	.993	1.269	1.360	1.472
54	7	3.56	8.2	.993	.930	1.174	.940	1.251	1.661	1.547
55	7	3.78	7.7	.987	.921	1.144	1.020	1.481	1.695	1.513
56	7	3.87	7.5	.986	.914	1.145	1.024	1.548	1.757	1.562
57	7	3.94	7.3	.990	.925	1.176	.943	1.372	1.728	1.741
58	7	4.28	6.7	.987	.920	1.177	.945	1.485	1.791	1.913
59	7	4.57	6.3	.984	.915	1.178	.947	1.588	1.851	2.066
60	7	4.63	6.2	.977	.892	1.180	1.015	1.873	2.079	2.086
64	7	6.67	4.4	.946	.811	1.280	1.009	3.160	3.516	3.404
86	3	2.61	11.9	1.002	1.008	.969	.854	.934	1.319	1.787
87	6	2.52	12.5	1.005	1.015	.977	.963	.818	.918	.849
88	7	2.70	11.4	1.003	1.004	.978	.974	.903	1.000	.899
89	7	3.17	9.4	.993	1.015	.990	.885	1.278	1.418	1.186
90	7	3.21	9.2	.998	.977	1.110	.941	1.073	1.203	1.487
91	7	3.22	9.2	.996	1.013	.975	.899	1.158	1.171	1.181
92	7	3.44	8.5	.993	.975	.994	.964	1.271	1.394	1.300
93	7	3.59	8.1	.993	.986	1.036	.947	1.249	1.222	1.507
94	7	3.61	8.1	.987	1.004	.977	.858	1.504	1.554	1.391
95	7	4.20	6.8	.983	.976	.969	.927	1.640	1.438	2.109
99	8	3.05	9.8	.999	1.004	.981	.958	1.029	1.057	1.156

Discussion

In this section we use $E(T_n)$ as shorthand notation for the "normalized" expectation $E(T_n / \sigma) / [E(T_n / \sigma)$ from normal distribution] and $C(T_n)$ for the normalized squared coefficient of variation $CV^2(T_n) / [CV^2(T_n)$ from normal distribution], where T represents S, M, or R as appropriate. Thus $E(R_n / d_n\sigma)$ is denoted by $E(R_n)$.

A number of attempts were made in trying to find properties of the parent distributions that could be used to summarize the variation in the expectation and coefficient of variation of S, M, and R. However, of the parent properties considered only k and β_2 seem to have any merit. Other properties considered were : β_1, δ/σ, ν_3 / σ^3, ν^5/σ^5 .

For any given value of β_2 it appears that the normalized expectation of the three estimators is considerably less for small values of k (say less than five or six) than it is for larger values of k even for the small sample sizes considered here. The normalized coefficient of variation, on the other hand, is considerably larger for small values of k. These very discrete distributions (with k less than five or six) would seem to be of little practical importance and since the properties of S, M, and R are substantially different for such distributions and would require separate treatment, only distributions with $k \geq 7$ will be considered further.

An interesting approach to describing the variation in the properties of S is due to Le Roux (1931). This method, which has been mentioned briefly in Section 4.3 above, is based on the use of a modified "degrees of freedom." The method suggests that the distribution of S based on a sample size n from a distribution with standardized fourth moment β_2 is approximately the same as that of an S from a sample of size n^* from a normal distribution, where $\nu = n-1$,

$$\nu^* = \nu \left[1 + \tfrac{1}{2} \frac{\nu}{\nu+1} (\beta_2 - 3)\right]^{-1}$$

and

$$n^* = \nu^* + 1 .$$

This approach gives $CV^2(S^2)$ exactly for all distributions and Le Roux has shown that the third and fourth moments of S^2 also agree quite well for samples from a subset of the Pearson family.

Table 4.5.4 gives values of $E(S_n)$ and $C(S_n)$ based on this modified degrees of freedom approach. A comparison between these approximate values and the results for $E(S_n)$ and $C(S_n)$ given in Table 4.5.3 shows good agreement for $k \geq 7$. The agreement seems to improve with increasing sample size. For $n=3$ and small values of β_2 (say 2 to 4) the values of $E(S)$ for discrete distributions tend to be slightly less than would be indicated by the approximation. Consequently, for the same values of β_2 , $C(S_3)$ tends to be slightly larger than the approximation.

Table 4.5.4

Approximate values for $E(S_n)$ and $C(S_n)$ for samples of size 3, 5, and 10 based on the "modified degrees of freedom" approximation.

	$E(S_n)$			$C(S_n)$		
β_2	n = 3	5	10	n = 3	5	10
2.0	1.0396	1.0248	1.0125	.6518	.5888	.5439
3.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4.0	.9637	.9763	.9877	1.3572	1.4227	1.4651
5.0	.9306	.9537	.9756	1.7208	1.8548	1.9383
6.0	.9003	.9322	.9638	2.0890	2.2942	2.4187
7.0	.8726	.9119	.9523	2.4607	2.7396	2.9057
8.0	.8470	.8927	.9410	2.8350	3.1899	3.3985
9.0	.8235	.8744	.9300	3.2115	3.6442	3.8965
10.0	.8017	.8571	.9193	3.5895	4.1019	4.3991

It is convenient to compare all of the results obtained here on S, M, and R with a single approximation. The modified degrees of freedom approximation was chosen as a standard since it gave very good results for S and consequently might give reasonably good results for M and R because of their high correlation with S in small samples.

For $n=3$ the three agree quite well with the approximation. The least satisfactory agreement is for $E(R)$ where the discrete results are about four percent lower on the average than the approximation. The results on $E(S)$ are about two percent lower on the average and those of $E(M)$ are about one percent lower. Most of the results on the coefficient of variation are slightly larger than the approximation for all three estimators.

For $n=5$ the $E(S)$ and $E(M)$ are in good agreement with the approximation except that the discrete results are on the average about one percent lower than the approximation. The $E(R)$ is again about four percent less than the approximation on the average. However, for $k=7$ the values of $E(M)$ and $E(R)$ seem to be less in agreement with the approximation than do those for the larger values of k. It appears that smaller values of k tend to give smaller values of $E(M)$ and $E(R)$ for any given value of β_2.

For $n = 10$ results are only available for $k \leq 8$ and there is essentially no agreement between $E(M)$, $E(R)$ and the approximation while the agreement between $E(S)$ and the approximation seems to have improved. There is still reasonably good agreement between $C(M)$, $C(R)$ and the approximation, but the agreement is not as good as for $n = 3$ and 5. All but three of the values for $C(M)$ are larger than the approximation and all but five of the values for $C(R)$ are larger than the approximation. The values for $C(M)$ seem to average about ten percent larger than the approximation while the values for $C(R)$ are smaller than the approximation for low values of β_2 and larger than the approximation for high values of β_2.

It seems that for samples of size 10 or more the individual natures of S, M and R begin to assert themselves. The modified degrees of freedom approximation, derived by equating variances of S^2, continues to behave well in providing approximations for the lower moments of S, but begins to fail for M and R.

The $E(M/\delta)$ does not seem to be any more stable than $E(M/\sigma)$ for the distributions considered here. The $E(M/\delta)$ tends to be larger than its normal theory value and increases with β_2, while $E(M/\sigma)$ is more often below its normal theory value and decreases with β_2.

Johnson (1958) has given values for E(M) for samples from a Pearson Type III distribution with parameter α. For $n = 3$ and 5, his values are about the same as the largest discrete results for the corresponding values of β_2. The value for $\alpha = 1$, corresponding to the exponential distribution with $\beta_2 = 9$, is about ten percent larger than the approximation for $n = 3$; about five percent larger for $n = 5$; and about one percent smaller for $n = 10$. In all cases the difference between the values for the type III distribution and the approximation decreases with increasing α (decreasing β_2) until complete agreement is attained for the limiting normal distribution.

In the same paper in which Le Roux suggested the modified degrees of freedom approximation to the distribution of S^2, he also reported some results on S^2 obtained from sampling studies. These results have been summarized in Section 4.3. Very good agreement exists between his results for $E(S_5)$ and $E(S_{10})$ and the approximation but for the larger value of β_2 his results for $C(S_5)$ and $C(S_{10})$ are considerably smaller than those given by the approximation.

The bounds given in Section 4.3 for E(S) appear to be quite conservative with respect to most of the distributions used here except for the distributions with very small values of k.

In Chapter 7 we make use of the information gained here on the behavior of the coefficient of variation of the estimators; in particular we use the modified degrees of freedom approach to compute the approximate effect of non-normality on four Cochran-type tests. We note here that for all three estimators the rate of change of the coefficients of variation with β_2 is roughly of the same order of magnitude as indicated by the Le Roux modified degrees of freedom approximation to the coefficient of variation of S. However, there is still considerable residual scatter for all three estimators and the quality of the approximation varies from estimator to estimator.

The approximation is best for S and seems to improve with n, while for M the approximation is almost as good but tends to underestimate $C(M)$ and not increase quite as rapidly with β_2 as it should.

For R the rate of change of $C(R)$ with β_2 is significantly underestimated by the approximation for these discrete distributions. Fortunately other results on R are available and comparisons made in the following section seem to indicate that R may be better behaved for other classes of distributions than it is for these discrete distributions.

For these discrete distributions there is considerably better agreement among the actual coefficients of variation of all three estimators, than there is with the approximation. Thus it is not possible to compare with any certitude a value for C(R) from a particular distribution with the approximate value of C(S) used here. There is some evidence that as the sample size and/or k increase the C(S) tends toward better agreement with the approximation.

4.6 SUMMARY OF RESULTS ON R

In this section we compare our results on the expectation and coefficient of variation of range in small samples from non-normal populations, with those available in the literature. This section may be viewed as an updating for $n = 3$, 5 and 10 of Cox's (1954) summary of these same properties. Throughout this section we use the notation E_n to denote $E(R_n/d_n\sigma)$ and C_n to denote $CV(R_n)/[CV(R_n)$ from normal distribution]. Thus, for the normal distribution E_n and C_n would be 1.0.

The comparisons made here are based on results that may be found in the following places:

1. Cox (1954)
2. Burr (1966)
3. Tukey random variables - Table 4.4.1 and 4.4.2
4. Miscellaneous exact results - Tables 4.3.5 and 4.3.6
5. Discrete distributions - Table 4.5.3
6. Gephart sampling study - Table 4.3.3

From Cox's (1954) results we use only his "average" values for E_n and C_n since there are no published tables for his individual points. Cox obtained his "average" by plotting a number of results on E_n and C_n as a function of β_2. He drew smooth curves through the points and read off "average" values as a function of β_2. This was done for $n = 2(1)5$. The values used here for E_3 were obtained by averaging his readings for E_2 and E_3 since the two should theoretically agree.

We use the terminology TRV-small and TRV-large to distinguish the Tukey random variable results for the smaller and larger values of λ, respectively. For more details see Section 4.4.

We consider only distributions having β_2 values in the interval $[2,9]$.

Comparisons on the Expectation of R

For the expected value of R there is good agreement between the Burr and TRV-small results for $n = 3$, while the miscellaneous results are very slightly lower on the average. Cox's "average" is roughly 1 or 2 units lower than the Burr and TRV-small results. The overall scatter of points from these four sources is about 2 units for any given value of β_2. For convenience these four sources will sometimes be referred to as "the four."

For $n = 5$ the Burr and miscellaneous results agree fairly well. The TRV-small results are about 1 unit higher than the above two for larger values of β_2 while Cox's "average" is about 1 or 2 units lower for very small values of β_2. The overall scatter of these points is again about 2 units for any given value of β_2.

For $n = 5$ the TRV-large results agree with "the four" for small values of β_2 (less than about $3\frac{1}{2}$ or 4) but for larger values of β_2 the TRV-large results drop off sharply. For $n = 3$, also, the TRV-large results drop off sharply with β_2, this time starting from $\beta_2 = 2$. For $\beta_2 = 6$ they are about 10 units lower than "the four" for $n = 3$, and about 5 units lower than "the four" for $n = 5$.

For $n = 10$ the patterns are not nearly so clear. Burr's values tend to lie above those for the miscellaneous distributions while the TRV-small results are somewhat larger than both for values of β_2 greater than 4. Cox's "average" is not available for $n = 10$. The TRV-large values are the largest overall and attain Plackett's (1947) maximum for $\beta_2 \doteq 4.66$ ($\nu = 9$). For a given value of β_2 the overall range of variation of the Burr, TRV-small and miscellaneous results is about 5 units.

There is considerable scatter to Gephart's results but on the average they tend to agree with the Cox, Burr, TRV-small and miscellaneous results for all three values of n. It is possible to compare Gephart's sampling results with exact values for 3 of his non-normal distributions or a total of 7 values of E_n. The sampling results are larger than the true value for 6 of the 7 cases.

In almost all cases the results for discrete distributions lie considerably below "the four," expecially for small values of k. For $n = 3$ and 5 there is virtually no overlap between the discrete results and the "Burr, miscellaneous, TRV-small" results. Some discrete results are larger than Cox's average for small values of β_2 (less than about 3) but there is no overlap for larger β_2. For $n = 10$ about one-third of the discrete results for distributions with $k = 7$ or 8 overlap the "Burr, miscellaneous, TRV-small" results. Overall the Cox, Burr, miscellaneous, and TRV-small results on E_n are slightly higher than those of Cox (1954). This may be because of the bimodality of many of the distributions used by Cox. We have noted here and in Section 4.5 that for a given value of β_2 the expectation of the range tends to decrease with increasing discreteness. Heuristically, it seems reasonable that a bimodal distribution should exhibit some of the tendencies of that extremely bimodal distribution whose probability is all concentrated at two points, that is, a discrete distribution with $k = 2$.

Comparisons on the Coefficient of Variation of R

For the coefficient of variation (denoted by C_n) there are very few results available for miscellaneous distributions. In most cases where results do exist they agree quite well with the TRV-small results. Hence, they will not be explicitly mentioned in the following discussion, but most of the commentary applicable to TRV-small results will also be applicable to the few miscellaneous distributions for which results are available.

It is well known that the expectation of the range is in general quite stable while the coefficient of variation of the range varies considerably from distribution to distribution. Consequently "good" agreement here will mean that the results on C_n differ by about 3 units or less implying that the squared coefficients of variation agree to within roughly 5 to 10 units, whereas for the expectation good agreement meant that the results on E_n differed by less than 1 unit.

For $n = 3$ the Burr results are about 2 to 5 units lower than the TRV-small results while Cox's average agrees with the TRV-small results to within about 2 units for β_2 less than about 4.5. For larger values of β_2 Cox's average gradually becomes larger than the TRV-small results until, at $\beta_2 = 8$, there is about a 10 unit discrepancy.

For $n = 5$ the Burr results agree quite nicely with the TRV-small results for β_2 less than about 4, while for larger β_2 the Burr values average about 5 units less than the TRV-small values. Cox's average is larger than the above two for β_2 greater than about 3 and becomes about 10 units larger than the TRV-small values for β_2 greater than 5.

For $n = 10$ there is good agreement between the Burr and TRV-small and large results for β_2 less than about 4. For larger values of β_2 the Burr results become smaller then the TRV-small results and the TRV-large results become larger, the difference being about 10 and 20 units respectively for $\beta_2 = 6$.

There is again considerable scatter in the Gephart empirical results but on the average they tend to agree with the Cox, TRV-small and Burr results.

The discrete results are almost universally larger than the Cox, TRV-small and Burr results especially for the smaller values of k. For the larger values of k the discrete results approach the above three but the values for $k = 11$ still average about 15 or 20 units larger for large values of β_2. For smaller values of β_2 the agreement for $k = 11$ is within about 5 units.

For the larger values of β_2 the Cox, TRV-small, Burr and miscellaneous results have C_n values that are slightly lower overall than Cox's "average" values, suggesting that for these distributions the coefficient of variation of R is slightly more stable than indicated by Cox's average. For the TRV-large and discrete distributions the coefficient of variation of R is considerably less stable than Cox's average.

4.7 MODIFIED SAMPLE SIZE APPROXIMATIONS FOR R

In Section 4.5 we have described Le Roux's method of approximating properties of an estimator for non-normal samples by its properties in normal samples of a modified size. We have also given some evidence that this method seems to work fairly well for the expectation and coefficient of variation of S. In this section we suggest a similar approach for R and give values for the modified sample size as a function of β_2 and n for several classes of distributions.

Le Roux's expression for the modified sample size n^* is simply

$$n^* - 1 = (n-1)\left[1 + \frac{1}{2}\,\frac{n-1}{n}\,(\beta_2 - 3)\right]^{-1}$$

where n is the sample size and β_2 is the standardized fourth moment of the parent distribution. This formula is such that the $CV(S^2)$ is given exactly. Here we continue to use the coefficient of variation as the basis for determining the modified sample size.

Table 4.7.1 gives values for the modified sample size n^* determined so that $CV(R)$ is given exactly. Three classes of distributions are included: TRV-small, TRV-large, and Cox'x "average." Thus, for example, a sample of size 5 from a TRV-small distribution with $\beta_2 = 4.2$ has the same $CV(R)$ as a sample of size 4 from the normal distribution. Other properties of R for samples of size 5 from a TRV-small distribution with $\beta_2 = 4.2$ should also be at least roughly the same as the corresponding properties of R for samples of size 4 from the normal distribution.

Table 4.7.2 gives similar results presented in a slightly different manner. Here values of β_2 are given corresponding to combinations of n and n^*. This table facilitates interpretation of the results of Section 7.2.

These tables essentially echo the discussion given earlier in this section and emphasize the finding that the behavior of the coefficient of variation of R, and consequently the modified sample size, depends considerably on the class of non-normal distributions considered. The CV(R) may be either more or less stable than $CV(S^2)$.

The three approximations for the distribution of R, corresponding to the TRV-small and TRV-large families and Cox's "average," will be used in Section 7.2 to describe the dependence of the robustness of the Bliss, Cochran, Tukey (1956) test on the class of non-normal distributions considered.

Table 4.7.1

Modified sample size necessary to make $CV^2(R)$ for normal distribution equal that for a sample of size n from indicated non-normal distribution. Also shown is Le Roux's modified sample size for S^2.

	S^2			Cox's "average"		TRV-small			TRV-large		
β_2	n=3	5	10	n=3	5	n=3	5	10	n=3	5	10
2	4.0	7.7	17.4	3.5	7.5	3.5	8	33	3.5	8	36
3	3.0	5.0	10.0	3^-	4.5^+	3	5	10	3^-	4.5	11^-
4.2	2.4	3.7	6.8	2.5	3.5^+	2.5^+	4	6.5	2.5^-	3^+	6.5
6	2.0	2.8	4.8	2.5^-	3	2.5	3.5	5	< 2	2.5	4
8	1.8	2.3	3.8	2^+	3^-	2.5^-	3^+	4.5	< 2	2	3

Table 4.7.2

Values of β_2 corresponding to the indicated modified sample sizes for R for three classes of distributions. Values based on the Le Roux modification for S^2 are also shown.

	S^2			Cox's "average"		TRV-small			TRV-large		
n^*	n=3	5	10	n=3	5	n=3	5	10	n=3	5	10
2	6.0	10.5	21	9.5	-	-	-	-	5.0	17.8	-
3	3.0	5.5	10.8	2.8	6.0	3.0	9.0		2.7	4.4	8.1
4	2.0	3.8	7.4		3.6		4.1	10.0		3.3	6.0
5	1.5	3.0	5.8		2.8		3.0	6.0		2.8	5.0
7	1.0	2.2	4.1		2.1		2.2	4.0		2.2	4.0
8		1.9	3.6		1.9		2.0	3.5		2.0	3.7
10		1.6	3.0		1.6			3.0			3.2
17		1.1	2.0					2.4			2.5

5. GENERAL PROPERTIES OF COCHRAN-TYPE TESTS

In this chapter we consider some general properties of tests of the null hypothesis

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 ,$$

of the form

$$Y = \max_{1 \le i \le k} V_i \Big/ \sum_{j=1}^{k} V_j$$

where the V_j are independent estimates of dispersion. The tests are such that H_0 is rejected if Y is significantly large.

In the first section we give a brief outline of the historical development of Cochran-type tests and in the second section we introduce some notation and give some general properties of the tests. In the third section we describe ways for obtaining approximations to the distributions of the tests.

5.1 HISTORICAL DEVELOPMENT

The following brief outline of the development of Cochran-type tests is not intended to be exhaustive, but is included in order to summarize the major stages in their development. Other works of interest will be cited at other points as the need arises.

Fisher (1929) gave the distribution of $Y = \max V_j \,/\, \sum_{j=1}^{k} V_j$ where the V_j are independently distributed as $\sigma^2 \chi_2^2 / 2$ where χ_ν^2 denotes a random variable distributed as chi-square with ν degrees of freedom. He proposed the test as a means for testing whether the largest amplitude in a harmonic analysis was significantly large. He included exact and approximate percent points for $\alpha = .05$ and $k = 5(5)50$, and later (1950) gave exact results for $\alpha = .05$, $.01$ and $k = 5(1)50$.

Cochran (1941) extended Fisher's results to the more general case where the V_j are independently distributed as $\sigma^2 \chi_\nu^2 / \nu$ for general ν and gave approximate 5 percent points for $\nu = 1(1)6(2)10$ and $k = 3(1)10$. He also noted that the method "is occasionally helpful in testing one of a group of estimates of variance which appears to be anomalously large." This latter statement appears to be the reason that the test bears his name rather than Fisher's.

Eisenhart and Solomon (1947) used the same methods as Fisher and Cochran to calculate approximate 1 and 5 percent points for $\nu = 1(1)10,16,36,144,\infty$ and $k = 2(1)10,12,15,20,24,30,40,60,120,\infty$.

Bliss, Cochran and Tukey (1956) introduced a test in which the V_j are independently distributed, each as the range of a sample of size n from a normal population. They gave approximate 5 percent points for $n = 2(1)10$ and $k = 2(1)10,12,15,20,50$.

Darling (1952) gave, in integral form, the characteristic function of $1/Y$ when the V_j have an arbitrary distribution. He also obtained the exact distribution of $1/Y$ when the V_j are distributed uniformly on $(0,a)$, and when the V_j are distributed as chi-square with even degrees of freedom.

Truax (1953) showed that when the V_j are distributed as $\sigma^2 \chi_\nu^2 / \nu$ the test based on Y is optimal in a decision theoretic sense. That is, under mild restrictions, this test maximizes the probability of the correct decision when the k+1 possible decisions are

$$D_0: \quad \sigma_i^2 = \sigma_j^2 \qquad \text{all } i, j$$

$$D_j: \quad \sigma_j^2 = \max(\sigma_1^2, \ldots, \sigma_k^2)$$

$$\text{and} \quad \sigma_j^2 \neq \sigma_i^2 \quad \text{for some } i.$$

Doornbos and Prins (1956) have given expressions for bounds on the power of Cochran's test and have also given a way of generalizing the test to include S^2's based on different degrees of freedom.

Siotani (1955) has compared two approximations to the percentage points of Cochran's test and has given some properties of an analagous test for the two largest variances.

5.2 MATHEMATICAL STRUCTURE OF THE TESTS

A Cochran-type test (CTT) has been defined to be a test based on the statistic $Y = \max V / \sum_{j=1}^{k} V_j$ where the V_j are independent estimates of dispersion and the null hypothesis,

$$H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 ,$$

is rejected when Y is significantly large.

We assume that the V_j are always positive and that they may be considered to be estimates of some power of σ_j . That is, if a linear transformation of the form

$$X^* = a + bX$$

is applied to the underlying observations, then the resulting V_j^*'s will be related to the V_j by

$$V_j^* = |b|^p V_j$$

where p does not depend on a or b. We also assume that under the null hypothesis the V_j are identically distributed and that under an alternative hypothesis the distributions are the same except for a scale factor which is a power of σ. Under these assumptions, Cochran-type tests are invariant when the same linear transformation is applied to the observations from each population. Hence, when considering the properties of Cochran-type tests we may assume that under the null hypothesis the scale parameters

σ_j are all unity and that under alternatives hypothesis one or more of the σ_j are different from unity.

The distribution of Y is usually intractable and it is usually most convenient to approximate it through the distribution of one of its components Y_i defined by

$$Y_i = V_i \Big/ \sum_{j=1}^{k} V_j .$$

It is desirable to have another quantity, say W_i, which is a monotonic transformation of Y_i but with independent numerator and denominator. Such a quantity is

$$W_i = V_i \Big/ \Sigma_i^* V_j$$

where

$$\Sigma_i^* V_j = \sum_{j=1}^{n} V_j - V_i .$$

Then

$$Y_i = (1 + 1/W_i)^{-1}$$

is a monotonic increasing transformation and a test on Y is equivalent to a test on W where $W = \max W_i$.

It is convenient to introduce the notation

$$U_i = V_i \Big/ B_i$$

where the B_i are appropriate divisors that make the U_i identically distributed. For the four tests considered in this study, B_i is proportional to σ_i^2 for the test based on S^2, and is proportional to σ_i for the other three tests.

Let the cumulative distribution function of the U_i be given by $F(u)$ for u in the interval $[0,\infty)$. Then the joint probability element of the U_i is

$$dF(u_1,u_2,\ldots u_k) = \prod_{j=1}^{k} dF(u_j) \qquad 0 \leq u_j < \infty \quad \text{for all } j.$$

The probability element of any one of the W_i may be expressed in integral form as

$$dG(w_i) = \underbrace{\int_0^\infty \cdots \int_0^\infty}_{(k-1)\text{fold}} (\Sigma_i^* u_j) dF(u_i \Sigma_i^* u_j) \; \Pi_i^* \; dF(u_j) \qquad 0 \leq w_i < \infty ,$$

where Σ_i^* has been defined above and Π_i^* is analogously defined to be

$$\Pi_i^* X_j = \prod_{j=1}^{k} X_j / X_i \quad .$$

When the U_i are distributed as chi-square, the distribution of W_i reduces to a scaled F-distribution, but in general the distribution of W_i does not have a simple form and the distribution of $W = \max(W_i)$ is usually even more complex. Ways to obtain approximate percentage points of the distribution of W from those of W_i will be discussed in Section 5.3.

The probability element of Y_i may be written as

$$dG(y_i) = (1-y_i)^{-2} \underbrace{\int_0^\infty \cdots \int_0^\infty}_{(k-1)\text{fold}} (\Sigma_i^* u_j) dF\left(\frac{y_i \Sigma_i^* u_j}{1-y_i}\right) \Pi_i^* dF(u_j) \qquad 0 \leq y_i \leq 1 .$$

This becomes a scaled beta when the U_i are distributed as χ^2. When the U_i are uniformly distributed on $(0,\theta)$ Darling (1952) has shown that the distribution of $1/Y$ is the same as that of $1 + X_1 + X_2 + \ldots + X_{k-1}$ where the X_j are uniformly distributed on $(0,1)$. No other non-trivial situation in which the distribution of Y_i or Y is simple, appears to be known.

5.3 COMPUTATION OF APPROXIMATE PERCENTAGE POINTS

In this section we show that for small values of k and α, percentage points of Cochran-type tests may be obtained without error from the distribution of W_i. For larger values of k and α, approximate mehtods such as those given by Fisher (1929) and Bliss, Cochran, and Tukey (1956), must be used. These two similar methods of approximation use percentage points of W_i to approximate percentage points of W. This seems to be the most convenient approach even when the distribution of W_i is not known exactly. The distribution of W_i may be obtained from those of its independent numerator and denominator by quadrature when they are known or when good approximations are available.

Proceeding more formally, let w_α be such that under the null hypothesis H_0,

$$P(W \geq w_\alpha) = \alpha ;$$

then this section is concerned with general methods for the determination of w_α. First it is convenient to introduce some more notation. Let

$$P(w) = P(W \geq w)$$

$$P_1(w) = P(W_i \geq w) \qquad \text{preselected } i$$

$$P_2(w) = P(W_i \geq w \ \text{ and } \ W_j \geq w) \qquad i \neq j \text{ preselected}$$

$$\vdots$$

$$P_k(w) = P(\text{all } W_i \geq w).$$

Now we prove the following theorem which was stated by Cochran (1941)

<u>THEOREM</u>:

$$P_2(w) = 0 \quad \text{if} \quad w > 1$$

and in general

$$P_\ell(w) = 0 \quad \text{if} \quad w > 1/(\ell-1) .$$

PROOF: Noting that

$$P_2(w) = P(W_j \geq w \mid W_i \geq w) \cdot P(W_i \geq w),$$

let $i=1$ and $j=2$ for definiteness. Then

$$\left\{ \frac{U_1}{U_2+U_3+\ldots+U_k} \geq w > 1 \right\} \Rightarrow \left\{ \frac{U_1}{U_2} > 1 \right\}$$

$$\Rightarrow \left\{ \frac{U_2}{U_1} < 1 \right\} \Rightarrow \left\{ \frac{U_2}{U_1+U_3+\ldots+U_k} < 1 < w \right\} .$$

To show $P_3(w) = 0$ if $w > \frac{1}{2}$ consider the following:

$$\left\{ \frac{U_1}{U_2+U_3+\ldots+U_k} > \tfrac{1}{2} \quad \text{and} \quad \frac{U_2}{U_1+U_3+\ldots+U_k} > \tfrac{1}{2} \right\}$$

$$\Rightarrow \left\{ \begin{array}{l} U_3 < 2U_1 - U_2 \\ U_3 < -U_1 + 2U_2 \end{array} \right\} \Rightarrow \left\{ 2U_3 < U_1 + U_2 \right\}$$

$$\Rightarrow \left\{ \frac{U_3}{U_1+U_2+U_4+\ldots U_k} < \tfrac{1}{2} \right\} .$$

The general argument is clear and holds for all $\ell = 2,\ldots,k$. This completes the proof.

It is true in general by Bonferroni's inequalities (see Feller 1957, David 1956, or Hume 1965) that

$$kP_1(w) \geq P(w) \geq kP_1(w) - \binom{k}{2} P_2(w) . \tag{5.3.1}$$

Hence when $w > 1$, $P_2(w) = 0$ by the preceding theorem, so

$$kP_1(w) \equiv P(w) .$$

It follows that for values of k and α such that $w_\alpha > 1$, percentage points of W may be obtained exactly from the distribution of W_1 by solving for w_α in the relation

$$P_1(w_\alpha) = \alpha/k . \tag{5.3.1a}$$

Thus for small values of k and α, percentage points of Cochran-type tests may be obtained without error if the distribution of W_i is known exactly. For larger values of k and α such that $w_\alpha < 1$, approximate methods must be used.

Fisher's Method

The first method of approximation for values of $w_\alpha \leq 1$, which is apparently due to Fisher (1929), is based on the use of equation (5.3.1a) even though it does not hold exactly. Fisher was interested in the 5 and 1 percent points of what is here called a "Cochran-type test" with the U_i distributed as a multiple of a chi-square with two degrees of freedom. He compared exact and approximate 5 percent points for this test and showed that the above method of approximation based on (5.3.1a) gave results that were accurate to within a unit in the fourth decimal for $k \leq 50$.

Letting w_α^* be such that

$$P_1(w_\alpha^*) = \alpha/k ,$$

it follows from equation (5.3.1) that

$$\alpha \geq P(w_\alpha^*) \geq \alpha - \binom{k}{2} P_2(w_\alpha^*) .$$

Thus when $P_2(w_\alpha^*)$ is small, Fisher's approximation w_α^* will be a good approximation to w_α .

Cochran (1941) used the same method of approximation in the computation of approximate 5 percent points for the test based on U_i distributed as chi-square with $\nu = 1(1)6(2)10$ degrees of freedom and $k = 3(1)10$. In the evaluation of the accuracy of the approximation, Cochran makes use of some exact results and apparently uses the inequality (which does not hold in complete generality)

(5.3.2) $$P_2(w) \le P_1^2(w) .$$

When this inequality holds, the following inequalities hold for all k:

$$.10 \ge P(w_{.10}^*) \ge .095$$

$$.05 \ge P(w_{.05}^*) \ge .04875$$

$$.01 \ge P(w_{.01}^*) \ge .00995 .$$

These bounds would seem to indicate that the approximation should be sufficient for many purposes.

It seems "intuitively obvious" that inequality (5.3.2) should hold for all $f(u)$, but a counterexample given by Kesten (see Doornbos 1956) has shown that this is not true. Doornbos, and Prins (1956) have, however, shown that the inequality does hold when the U_i are distributed as chi-square (with arbitrary degrees of freedom).

The counterexample given by Kesten may be outlined as follows : Let $k = 3$ and U_1, U_2, U_3, independently take on the values 1 and 2 with probability p and $q = 1-p$ respectively. Then for values of w in the interval $(\frac{1}{2}, \frac{2}{3}]$, $P_1(w) = p^2q + 2pq^2$ and

$P_2(w) = pq^2$ so for q in the interval $\frac{1}{2}(\sqrt{5} - 1) < q < 1$, it follows that $P_2(w) > P_1^2(w)$ contrary to (5.3.2). For such values of w and q, $P(w) = 1 - p^3 - q^3$ so the counterexample can be shown to hold for any α in the interval $0 < \alpha < 3\sqrt{5} - 6 \doteq 0.7083$ for appropriately selected p and with $w_\alpha = \frac{2}{3}$. The interval on w originally given by Kesten seems to correspond to the interval $(\frac{1}{3}, \frac{1}{2}]$ in the present notation. However, the interval $(\frac{1}{2}, \frac{2}{3}]$ is preferable here since $\alpha=1$ for the former.

Bliss, Cochran, Tukey Method

A second method of approximation has been used by Bliss, Cochran, and Tukey (1956) in their tabulation of approximate percentage points of the ratio $R_{max} / \Sigma R_j$ where the R's are ranges of samples of size n from normal populations. If we let $P_1(w_\alpha) = \alpha'$, then their method is based on the following inequality (which also does not hold in complete generality):

$$(5.3.3) \qquad \frac{\alpha}{k} \leq \alpha' \leq 1 - (1-\alpha)^{1/k} \quad .$$

Bliss, et al. used a weighted average of the two bounds as a basis for their approximation. Let

$$(5.3.4) \qquad \alpha^0 = \frac{1}{3}\left[2(\alpha/k) + 1 - (1-\alpha)^{1/k} \right],$$

then α^0 is "approximately" equal to α' since (when the inequality 5.3.3 holds) they are bounded by the same quantities. Defining

w_{α^o} to be such that $P_1(w_{\alpha^o}) = \alpha^0$, yields w_{α^o} as an approximation to w_α.

Some examples of the ratio of the bounds in this inequality are listed in Table 5.3.1 for α = .10, .05, and .01 and several values of k. It would seem that when the inequality holds, these bounds would be satisfactory for most purposes. The left inequality in (5.3.3) follows immediately from (5.3.1) but the right inequality is not true in complete generality. Kesten's counterexample to the assertion of Cochran (1941) may be used as a counterexample here as well.

Table 5.3.1

Ratio of the bounds given by inequality (5.3.3).

$k \backslash \alpha$	.10	.05	.01
2	.974	.987	.997
5	.959	.980	.996
10	.954	.977	.995
20	.952	.976	.995
100	.950	.975	.995

COUNTEREXAMPLE:

Let U_1, U_2 and U_3 be defined as in the previous counterexample and let p assume some value in the interval (0, .5218). Then for $\alpha = 1-p^3 - q^3$, $w_\alpha = \frac{2}{3}$ and $1 - \alpha' = p + q^3$ making $(1 - \alpha) > (1 - \alpha')^k$ contrary to (5.3.3).

Approximate percentage points may be computed by the Bliss-Cochran-Tukey method in any case but there is at present no easily computable bound on the errors unless the inequality (5.3.3) can be shown to hold. Conditions under which the right inequality of (5.3.3) holds, would seem to be similar to those under which the inequality (5.3.2) would hold.

Lehmann (1966) has recently discussed such inequalities in some generality and has termed inequalities of the type (5.3.2) "negative quadrant dependence." He suggests that the chain of conditions:

(I) $P(W_2 \leq w_2 \mid W_1 = w_1)$ is non-decreasing in w_1,

(II) $P(W_2 \leq w_2 \mid W_1 \leq w_1) \leq P(W_2 \leq w_2 \mid W_1 \leq w_1')$ for all $w_1 < w_1'$ and all w_2

(III) $P(W_2 \leq w_2 \mid W_1 \leq w_1) \leq P(W_2 \leq w_2)$ for all w_1, w_2,

may sometimes be used to prove negative quadrant dependence (Condition III is equivalent to inequality 5.3.2). In the above chain the conditions are linked by, Condition I implies Condition II, which implies Concicion III.

Lehmann used this chain of inequalities to obtain a simple proof that (5.3.2) holds for the case where the U_i are distributed as χ^2 by using Condition I and an explicit formula for the joint distribution of Y_i and Y_j. However, all three of the conditions are difficult to prove for most of the possible Cochran-type tests, and a general theorem would be useful. It would seem that a simple condition such as unimodality of the distribution of the U_i should be sufficient, but a number of such approaches have been tried without success.

Distribution of W_i Not Known

Sometimes the distribution of W_i is not known exactly, though the distributions of its component numerator and denominator are. When this is the case it may be convenient to integrate the following relation by quadrature:

$$(5.3.5) \qquad P(W_i \geq w) = \int_0^\infty G(y/w)dF(y)$$

or equivalently

$$(5.3.5a) \qquad P(W_i \geq w) = \int_0^1 \frac{G(-\log t/w)}{t} dF(-\log t)$$

where $G(y)$ is the c.d.f. of the denominator of W_i and $dF(y)$ is the probability element of the numerator of W_i.

Quadrature may also be convenient when the distributions of the numerator and/or denominator are not known exactly, but good approximations are available.

Error Analysis

There are in general two possible sources of error in the computation of approximate percentage points of CTT by the methods discussed in this section. Errors are introduced by (1) the approximation of w_α by w_α^* or $w_\alpha o$, and (2) by imperfect knowledge of the distribution of W_i. These two sources of error tend to be relatively unimportant from a practical point of view, compared to the general sensitivity of the tests to non-normality. A more detailed discussion of this third point is given in Section 7.2. Here we restrict attention to an analysis of the relative contribution of the first two types of error. An error analysis for the two new tests is given along with their computed percentage points in Chapter 6.

Fisher (1929), Cochran (1941), and Eisenhart and Solomon (1947) have used Fisher's method in the tabulation of approximate percent points of the test based on S^2's from a normal population. The distribution of W_i is well known in this case being a simple transformation of the F or incomplete beta distributions, both of which have been extensively tabulated. An error analysis by Cochran indicated that the approximation based on w_α^* gave values for w_α that were correct to four decimal places for $\alpha = .05$ and ν and k less than 10. For $\alpha = .01$ the results should be even more accurate.

For $\nu = 2$, a comparison is possible between some of the approximate results given by Eisenhart and Solomon (1947) using "Fisher's method" and the exact results given by Fisher (1950). This comparison showed that the tables of Eisenhart and Solomon had an error in the fourth decimal place of: 1 for $\alpha = .05$ and $k = 20$ and 24; 2 for $\alpha = .05$ and $k = 30$ and 40; and 1 for $\alpha = .01$ and $k = 40$. The $k = 5$, $\alpha = .01$ value of Fisher should be .78853 rather than .78874 as printed.

The conclusion drawn here is that the error introduced by the approximation of w_α by w_α^* or w_α^o is at most about half a unit in the third decimal place for $\alpha = .05$ and less than that for $\alpha = .01$. In three of the four CTT considered here the uncertainty in the computation of w_α^* or w_α^o is enough larger as to make the former source of error relatively unimportant.

A spot check of Eisenhart and Solomon's table indicated an occasional difference of 2 in the fourth decimal between the values tabled and the correct values of the <u>approximation</u>. A similar spot check indicated that the BCT approximation gives values, for Cochran's test, that differ from the values given by Fisher's approximation by at most 4 in the fourth decimal.

Harter's (1963) tables of percent points of the ratio of two ranges makes possible the computation of more accurate values of Bliss, Cochran, and Tukey's statistic for $k = 2$. These results are given in Table 5.3.2.

Table 5.3.2

Corrected values for $k = 2$ in Bliss, Cochran, Tukey (1956) table obtained from Harter's (1963) tables.

n	4	5	6	7	8	9	10
Corrected Value	.799	.759	.733	.714	.699	.688	.678

6. APPROXIMATE PERCENT POINTS OF TWO NEW TESTS

In this chapter approximate percent points of two new Cochran-type tests are computed. The first of the two tests was suggested by Churchill Eisenhart and is based on the standard deviation. The second test is believed to be a new one and is based on the mean deviation.

The primary motivation for suggesting these two new tests has been to obtain a test that is less sensitive to non-normality than those currently available. Some properties of these tests are given in Chapter 7 and approximate results obtained there indicate that there is little difference between the robustness of these tests and the tests of Cochran (1941) and Bliss, Cochran, and Tukey (1956). However, these tests may occasionally be useful for other reasons.

6.1 EISENHART'S TEST

Eisenhart's test is the Cochran-type test based on the statistic

$$E = \max S_i \Big/ \sum_{j=1}^{k} S_j$$

where the S_j are standard deviations, each having the same number of degrees of freedom and each based on an independent sample. Churchill Eisenhart has suggested this test in an attempt to obtain a more robust test than those currently available, and also because it may sometimes be more readily computable.

The reasons behind his conjecture that this test may be more robust than Cochran's may be outlined as follows : the square root of Cochran's test statistic is proportional to

$$\max S/S_p$$

where S_p denotes the pooled standard deviation defined by

$$S_p^2 = \sum_{j=1}^{k} S_j^2 / k \ .$$

Eisenhart's statistic analogously is proportional to

$$\max S/\bar{S}$$

where

$$\bar{S} = \sum_{j=1}^{k} S_j / k \ .$$

Written in this form, it is apparent that the two differ only in the manner in which the information in the denominator is combined. Eisenhart reasoned that $\bar{S}$ would be "less sensitive" to non-normality than S_p so the test based on $\bar{S}$ should be more robust. A more detailed comparison of the tests is given in Chapter 7 below.

In this section we are concerned with the tabulation of approximate percent points of E under the null hypothesis H_0 and under the basic assumption that each sample is independently drawn from a normal distribution. We use the Bliss-Cochran-Tukey method of obtaining approximate percentage points of E from those of a component E_i defined by

$$E_i = S_i^2 \,/\, \sum_{j=1}^{k} S_j^2 \ .$$

This method of approximation has been discussed in Section 5.3 above.

Let y_α be such that (under the above assumptions)

$$P(E \geq y_\alpha) = \alpha$$

and let

$$W_i = E_i \,/\, (1-E_i) = S_i \,/\, \Sigma_i^* S_j \ .$$

Let w_p be such that

$$P(W_i \geq w_p) = p \ .$$

Then it has been shown in Section 5.3 that when $y_\alpha > \frac{1}{2}$,

$$y_\alpha = w_p \ / \ (1+w_p)$$

where

$$p = \alpha/k \ .$$

For cases where $y_\alpha \leq \frac{1}{2}$, let α^0 be defined by equation (5.3.4) then it has also been shown in Section 5.3 that

$$y_\alpha \doteq w_\alpha \ / \ (1 + w_{\alpha^0}) \ .$$

We now make use of the results obtained in Appendix A which show that the distribution of

$$T_i = \Sigma_i^* S_j$$

can be : (1) evaluated exactly in certain cases, (2) approximated quite well by the distribution of a fractional powered scaled chi with the same first three moments (Cadwell approximation), and (3) reasonably well approximated by a scaled chi with the same first two moments. We first compute approximate values of y_α from the chi approximations then evaluate the accuracy of these where k and ν are such that exact results are available. We then show that the Cadwell approximation agrees quite well with the exact values and use it to obtain slightly more accurate values of y_α for other values of m and ν where the chi approximations are not sufficiently accurate.

Let the distribution T_i be approximated by the distribution of

$$(K \chi_\xi^2 / \xi)^{\frac{1}{2}} ,$$

where χ_η denotes a random variable distributed as chi, the positive square root of a chi-square variate with η degrees of freedom. Then the distribution of W_i may be approximated by that of

$$(\chi_\nu / \sqrt{\nu}) / (K \chi_\xi^2 / \xi)^{\frac{1}{2}},$$

where ν is the degrees of freedom associated with each S_j. Since the ratio of two independent chi variates is proportional to the square root of an F variate, the approximate distribution of W_i may be expressed as a simple transformation of the incomplete beta distribution. Let $\beta_p(a,b)$ be the p'th fractile of the incomplete beta distribution with parameters a and b. That is, let $\beta_p(a,b)$ be defined by

$$(6.1.1) \qquad p = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} \int_0^{\beta_p(a,b)} t^{a-1} (1-t)^{b-1} \, dt .$$

Then $w_p \doteq \sqrt{(1-A) / (\nu KA)}$ where

$$A = \beta_p(\tfrac{1}{2}\xi, \tfrac{1}{2}\nu) .$$

The goodness of this approximation depends on the accuracy of the approximation to the distribution of $\Sigma_i^* S_j$ by that of a scaled chi. Approximate values for y_α were computed using the chi(T) approximation (defined in Appendix A) for $(k-1)\nu$ less than 30 and the chi(T^2) approximation for larger values of k and ν.

When the exact distribution of T_i can be expressed in simple form, the true probability associated with a given value of w_p can be obtained by quadrature using the relation in equation (5.3.5) or (5.3.5a). This was done for $\nu=1$ with $k=3,4$ and for $k=3$ with $\nu=2,4,6,8$ by using quadrature on (5.3.5a). The interval (0,1) was cut into 4 equal subintervals and 7 point Gaussian quadrature was used on each. This procedure was sufficient to give the required accuracy.

The Cadwell approximation can also be implemented through quadrature using the relation (5.3.5a). Let the distribution of T_i be approximated by that of $H\chi_\gamma^\lambda$ where H, γ, and λ are defined in Appendix A. Also let

$$P_a(x^2) = \int_0^{x^2} \frac{y^{\frac{1}{2}a-1} \; e^{-\frac{1}{2}y}}{\Gamma(\frac{1}{2}a) \; 2^{\frac{1}{2}a}} \, dy$$

denote the cumulative distribution of chi-square with a degrees of freedom. Then for the Cadwell approximation (5.3.5a) becomes

$$(6.1.2) \qquad p = \int_0^1 P_\gamma \left(2(-\log x/wH)^{1/\lambda}\right) g_\nu(-\log x)dx$$

where w is a trial value for w_α and p denotes the Cadwell approximation to the probability associated with w. Here $g_\nu(x)$

denotes the distribution of $\chi_\nu / \sqrt{\nu}$ given by

$$g_\nu(x) = 2(\frac{\nu}{2})^{\nu/2} x^{\nu-1} e^{-\frac{1}{2}\nu x^2} / \Gamma(\frac{\nu}{2}) .$$

Seven point Gaussian quadrature on 4 equal subintervals was again found sufficient to give the integral (6.1.2) to the required accuracy. The results of this computation are given in Tables 6.1.1 and 6.1.2.

Bliss, Cochran, and Tukey (1956) have shown for $\nu=1$ that it is also possible to compute percentage points of the range test (and therefore for the standard deviation test) from Lord's (1947) table. The results obtained by the "exact" and "Lord" procedures are compared with those based on the "Cadwell" approximation in Table 6.1.1. Where a comparison is possible, it appears that the values based on the Cadwell approximation are slightly more accurate overall than those obtained by interpolation in Lord's table. In only three cases is the value given by the Cadwell approximation in error by as much as 2 units in the third decimal place. Evidence has been given in Appendix A that the region covered by Table 6.1.1 is the region in which the Cadwell approximation is least accurate, so it is believed that the Cadwell approximation gives values that are accurate to within 2 units in the third decimal over the entire table.

Table 6.1.2 gives the differences between the chi and Cadwell procedures. The chi approximation does not give satisfactory accuracy for $\nu=1$ and k less than 15 or for $\nu=2$ and k less than 6 or 7. However, in other regions of the table the chi procedures appear to be accurate to within 1 or occasionally 2 units in the third decimal, which is considered sufficient accuracy for present purposes.

Table 6.1.3 gives approximate values for the percentage points of Eisenhart's test obtained from a combination of these three procedures. The exact values have been used for $k=3$ with $\nu=1,2,4$ and $k=4$ with $\nu=1$. The values given by the Cadwell approximation have been used for $\nu=1$ with $k=5,6,7,8,9,10,12,15$ and for $\nu=2$ with $k=4,5$. In all other cases the values obtained from the chi approximation have been used. The use of results obtained by several different methods makes the accuracy of the table uneven, but enables the most accurate value available to be used for each entry.

Bounds on the error introduced in y_α due to the approximation of α' by α^0 (when $y_\alpha < \frac{1}{2}$) can also be given if it can be assumed that the inequality (5.3.3) holds. Values for y_α were computed from α^0 and from the two extremes of the inequality. If the inequality holds, the error in the third decimal of y_α is at most : 2 for $\alpha = .10$, 1 for $\alpha = .05$ and 0.3 for $\alpha = .01$.

The difference between y_α and the supposed bounds is greatest for small values of ν and k such that y_α is slightly less than 0.5.

The errors introduced in the values reported in Table 6.1.3 due to the approximation of α' by α^0 apparently are greatest for $\alpha = .10$, while those introduced thru imperfect knowledge of the distribution of the denominator of W_i are greatest for $\alpha = .01$. We conclude that the combined error in Table 6.1.3 is less than a unit in the third decimal over much of the table, but for small values of ν and k some values may be in error by as much as 3 units in the third decimal.

The percent points of the incomplete beta distribution and the probability integral of the chi-square distribution were computed using the programs of Bargman and Ghosh (1963).

Table 6.1.1

A comparison of approximations to the percentage points of Eisenhart's test using the "Cadwell," "exact," and "Lord" procedures.

		$\alpha = .01$			.05			.10		
k	ν	C	E	L	C	E	L	C	E	L
3	1	.905	.907	.908	.810	.812	.813	.750	.751	.751
4	1	.786	.788	.784	.679	.679	.681	.621	.621	.622
5	1	.683		.681	.581		.581	.530		.530
6	1	.600		.599	.507		.508	.462		.462
7	1	.533		.535				.410		.410
10	1	.399		.398						
3	2	.754	.754		.663	.663		.616	.616	
3	4	.621	.621		.557	.557		.524	.524	
3	6	.564	.564		.512	.512		.486	.486	
3	8	.530	.530		.486	.486		.464	.464	

Table 6.1.2

Difference between the "Chi" and "Cadwell" approximations to the percentage points of Eisenhart's test. Entries are (Chi-Cadwell)X1000 and are for α = .01, .05 and .10 respectively.

k	$\nu = 1$	2	3	4	5	6	8	9
3	8 8 6	4 3 1	2 1 0	2 0 1	1 0 0	0 0 0	0 0 0	---
4	13 8 6	4 2 2	2 1 1	1 1 0				
5	12 7 4	3 1 1						
6	10 6 4	2 1 1						
7	9 4 3	1 1 0	1	0		0		0
8	6 3 3							
9	6 3 2							
10	4 3 1							
15	2 1 1							

Table 6.1.3

Approximate percentage points of Eisenhart's test $\max S_j \,/\, \sum_{j=1}^{k} S_j$

where each S is based on ν degrees of freedom.

(a) $\alpha = .01$

k	$\nu = 1$	2	3	4	5	6	7	8	9	11	14	19	29	49	99
2	.992	.934	.873	.828	.794	.769	.749	.732	.719	.698	.675	.649	.621	.592	.565
3	.907	.754	.671	.621	.588	.564	.545	.530	.518	.499	.479	.458	.433	.409	.387
4	.788	.615	.538	.493	.465	.444	.428	.415	.404	.388	.372	.353	.333	.313	.294
5	.683	.517	.449	.410	.384	.366	.352	.341	.332	.318	.304	.288	.270	.254	.237
6	.600	.446	.385	.350	.327	.311	.299	.290	.282	.270	.257	.243	.228	.213	.199
7	.533	.392	.336	.305	.285	.271	.260	.252	.245	.234	.223	.211	.197	.184	.172
8	.480	.351	.299	.271	.253	.240	.231	.223	.217	.207	.197	.186	.174	.162	.151
9	.435	.316	.269	.244	.228	.216	.207	.200	.194	.186	.176	.167	.155	.145	.135
10	.399	.288	.245	.222	.207	.196	.188	.182	.176	.168	.160	.151	.141	.131	.121
12	.343	.245	.208	.188	.175	.166	.159	.153	.149	.142	.135	.127	.118	.110	.102
15	.281	.200	.170	.153	.142	.135	.129	.124	.121	.115	.109	.103	.095	.088	.082
20	.218	.154	.130	.117	.109	.103	.099	.095	.092	.088	.083	.078	.072	.067	.062
50	.094	.066	.055	.050	.046	.043	.041	.040	.039	.037	.035	.032	.030	.027	.025

Table 6.1.3 (Continued)

(b) $\alpha = .05$

k	$\nu = 1$	2	3	4	5	6	7	8	9	11	14	19	29	49	99
2	.962	.862	.797	.756	.728	.707	.691	.678	.667	.651	.633	.614	.592	.570	.549
3	.812	.663	.596	.557	.531	.512	.498	.486	.477	.462	.447	.430	.411	.393	.375
4	.679	.535	.474	.440	.418	.402	.389	.379	.371	.359	.346	.332	.316	.300	.285
5	.581	.449	.396	.365	.345	.331	.320	.312	.305	.294	.283	.270	.256	.243	.230
6	.507	.387	.339	.312	.295	.282	.273	.265	.259	.249	.239	.228	.216	.204	.193
7	.450	.340	.297	.273	.257	.246	.236	.231	.225	.217	.208	.198	.187	.177	.166
8	.405	.304	.265	.243	.229	.218	.211	.204	.199	.192	.184	.175	.165	.155	.146
9	.368	.275	.239	.219	.206	.196	.189	.184	.179	.172	.165	.157	.147	.139	.130
10	.337	.251	.218	.199	.187	.178	.172	.167	.163	.156	.149	.142	.133	.125	.118
12	.290	.214	.185	.169	.159	.151	.146	.141	.137	.132	.126	.119	.112	.105	.099
15	.239	.176	.152	.138	.129	.123	.118	.115	.112	.107	.102	.097	.091	.085	.079
20	.186	.136	.117	.106	.099	.094	.091	.088	.085	.082	.078	.074	.069	.064	.060
50	.082	.059	.050	.045	.042	.040	.038	.037	.036	.034	.033	.031	.028	.026	.024

Table 6.1.3

(c) $\alpha = .10$

k	$\nu = 1$	2	3	4	5	6	7	8	9	11	14	19	29	49	99
2	.927	.813	.753	.717	.692	.674	.661	.650	.641	.627	.612	.596	.577	.559	.541
3	.751	.616	.558	.524	.502	.486	.474	.464	.456	.444	.430	.416	.400	.384	.369
4	.621	.495	.443	.414	.395	.381	.371	.362	.355	.345	.333	.321	.307	.294	.280
5	.530	.415	.370	.344	.327	.314	.305	.298	.292	.282	.272	.262	.249	.238	.226
6	.462	.358	.317	.294	.279	.268	.260	.253	.248	.240	.231	.221	.210	.200	.190
7	.410	.316	.278	.257	.244	.234	.227	.221	.216	.208	.200	.192	.182	.173	.164
8	.369	.283	.248	.229	.217	.208	.201	.195	.191	.184	.177	.169	.161	.152	.144
9	.336	.256	.224	.207	.195	.187	.181	.176	.172	.165	.159	.152	.144	.136	.128
10	.309	.234	.205	.188	.178	.170	.164	.160	.156	.150	.144	.137	.130	.123	.116
12	.267	.200	.174	.160	.151	.144	.139	.135	.132	.127	.122	.116	.109	.103	.097
15	.221	.165	.143	.131	.123	.118	.113	.110	.107	.103	.099	.094	.088	.083	.078
20	.173	.128	.110	.101	.095	.090	.087	.084	.082	.079	.075	.072	.067	.063	.059
50	.077	.056	.048	.043	.041	.039	.037	.036	.035	.033	.032	.030	.028	.026	.024

6.2 The Mean Deviation Test

The mean deviation test is based on the statistic

$$Z = \max M_i \Big/ \sum_{j=1}^{k} M_j$$

where the M's are defined by

$$M = \sum_{i=1}^{n} |X_i - \bar{X}|/n \ .$$

This test is proposed since some (e.g. Tukey 1960) believe that the properties of the mean deviation are less affected by departures from normality (especially in the direction of heavier tails) than those of either the standard deviation or the range. The effect of non-normality on the expectation and coefficient of variation of the mean deviation is considered in some detail in Chapter 4, and the effect of non-normality on the Cochran-type test proposed here is considered in Chapter 7 below.

The use of the mean deviation in tests for homogeneity of variance has been suggested by Cadwell (1953 and 1954). He has given approximate percentage points of a Hartley (1950) type test based on $\max M / \min M$, and has also given a procedure by which Bartlett's (1937) test can be applied to mean deviations.

In this section we compute approximate percent points of the mean deviation test under the null hypothesis H_0, and under the basic assumption that each sample is independently drawn from a normal distribution. More formally, let z_α be such that

$$P(Z \geq z_\alpha \mid H_0) = \alpha ,$$

then we are concerned here with the determination of approximate values for z_α. Let

$$Z_i = M_i \Big/ \sum_{j=1}^{k} M_j$$

and

$$W_i = M_i / \Sigma_i^* M_j = M/M^*$$

where

$$M^* = \Sigma_i^* M_j .$$

The computation of approximate percent points of this test is much more difficult than for Eisenhart's test since here it is necessary to approximate the distributions of both the numerator and denominator of W_i, whereas in Eisenhart's test the distribution of the numerator was known. Cadwell (1954) has, however, shown that the distributions of M and M^* can be reasonably well approximated by those of scaled chi-squares raised to the 1/(1.8) power and having the same first two moments. That is, the distributions of M and M^* are approximately the same as those of random variables distributed as

$$K(\chi^2_{\xi})^{1/(1.8)}$$

where K and ξ are chosen so that the approximations have correct mean and variance.

Let the distribution of M be approximated by that of $K_1(\chi^2_{\xi_1})^{1/(1.8)}$, and the distribution of M^* by that of $K_2(\chi^2_{\xi_2})^{1/(1.8)}$. Then the distribution of W_i may be approximated by that of $(K_1/K_2) (\chi^2_{\xi_1} / \chi^2_{\xi_2})^{1/(1.8)}$, which is a simple transformation of the incomplete beta distribution.

Let $A = \beta_p(a,b)$ be the p'th fractile of the incomplete beta distribution (defined in equation 6.1.1) with parameters $a = \xi_1/2$ and $b = \xi_2/2$

$$\text{where} \quad p = \alpha/k \quad \text{if} \quad z_\alpha \geq \tfrac{1}{2} \quad \text{and}$$
$$= \alpha^0 \quad \text{if} \quad z_\alpha < \tfrac{1}{2} \,,$$

and α^0 is defined by equation (5.3.4).

Then

$$z_\alpha \doteq D/(D + 1)$$

where

$$D = (K_1 / K_2) [(1-A) / A]^{1/(1.8)} \,.$$

Values of z_α for $k = 2(1)10$, $n = 4(1)10$ and $\alpha = .05$ were computed using this procedure. The values of K and ξ used were computed by the method outlined by Cadwell (1954). He has given 3 and 4 significant figure tables of ξ and K, respectively, but it was felt that more accurate values might improve the approximation. The results of these computations for z_α are given in Table 6.2.1.

The percentage points of the incomplete beta distribution were computed using the computer programs of Bargman and Ghosh (1963).

The error introduced in z_α due to the approximation of α' by α^0 should be of the same order of magnitude as in Eisenhart's test where the error is believed to be at most 1 in the third decimal place for $\alpha = .05$.

An assessment of the error introduced through imperfect knowledge of the distributions of the numerator and denominator of W_i is much more difficult. An investigation by Cadwell (1954) indicated that the approximation was in error by at most 1.5 in the third decimal in the same sense as discussed in Appendix A of this work. Since a maximum error of half this amount in the approximation to the denominator of Eisenhart's test produced an error of at most 2 units in the third decimal place in the Cadwell approximation to y_α, one is led to conjecture that the error in z_α caused by the imperfect knowledge of the denominator of W_i is at most 4 units in the third decimal.

Table 6.2.1

Approximate percentage points for the mean deviation test

for homogeneity of variance, $\max M_j \Big/ \sum_{j=1}^{k} M_j$.

Each M is based on n independent observations

from the normal distribution.

$\alpha = .05$

	n = 2	3	4	5	6	7	8	9	10
k = 2	.962	.862	.798	.760	.733	.713	.697	.685	.674
3	.812	.667	.600	.563	.537	.519	.505	.493	.484
4	.679	.538	.479	.446	.424	.408	.396	.386	.378
5	.581	.451	.400	.371	.352	.337	.327	.318	.311
6	.507	.389	.344	.318	.300	.288	.278	.271	.264
7	.450	.342	.302	.278	.263	.251	.243	.236	.230
8	.405	.305	.269	.248	.234	.223	.215	.209	.204
9	.368	.276	.243	.223	.210	.201	.194	.188	.183
10	.337	.253	.222	.204	.192	.183	.176	.171	.167

As a partial check on the accuracy of the present approximation, we computed values for z_α for $n=3$ using Cadwell's (1954) approximation. Table 6.2.2 gives the results of this computation and also gives the values of the approximation to the percent point of the range test computed by Bliss, Cochran, and Tukey (1956).

Table 6.2.2

Approximate values of $z_{.05}$ for $n=3$ from Bliss, Cochran, Tukey (1956) table and from method used here based on Cadwell's (1954) approximations.

k	3	4	5	6	7	8	9	10
BCT	.667	.538	.451	.389	.342	.305	.276	.253
Cadwell	.663	.537	.451	.389	.343	.307	.278	.254

Since it has been shown in Section 4.2 that the distributions of M and R are identical (except for a scale factor) for samples of size 3 from a normal distribution, the amount of disagreement between the two approximations should give some indication of the error in the methods. The two methods agree to within 2 units in the third decimal place for $k = 4(1)10$ while for $k=3$ they differ by 4 units in the third decimal place. The values of z_α should be somewhat more accurate for larger n since Cadwell (1954) indicated that the approximation used here gives reasonable accuracy for values of n greater than 3, but does not recommend its usage for $n = 3$.

In Table 6.2.1 the Bliss, Cochran, Tukey values have been used for $n = 3$ while for $n = 2$ the values obtained in Section 6.1 for Eisenhart's test have been used since in this case the tests based on the standard deviation, mean deviation and range are identical.

We believe that the values of z_α given in Table 6.2.1 should certainly be accurate to within 1 unit in the second decimal place and are probably accurate to within 5 units in the third decimal.

7. SOME SPECIFIC PROPERTIES OF THE FOUR TESTS

This chapter consists of two sections, the first of which constitutes an investigation of some power-type properties of the four Cochran-type tests under the basic assumption that all samples are independently drawn from normal distributions. The second section is concerned with the determination of some properties of the four tests when the samples are independently drawn from identical but non-normal distribution. Much of the notation used here has been introduced in Section 5.2.

In Section 7.1 we prove that for samples of equal size from normal distributions, Cochran's test is the likelihood ratio test of H_0 against alternatives of a certain family H_θ. We consider only this family of alternatives and show that for samples of size 3 and 5 there is little difference in the power-type properties of the four Cochran-type tests. However, for $n = 10$, Cochran's and Eisenhart's tests do seem to have a slight edge over the BCT and mean deviation tests in this respect.

In Section 7.2 we give evidence that the four tests under consideration here (and a few others) appear to have somewhat different robustness properties for samples from various non-normal distributions. None of the tests appears to be significantly more robust overall than the others, although this possibility has not been ruled out. However all of the tests considered are sufficiently non-robust to make the differences among the tests a second order consideration.

7.1 PROPERTIES UNDER NORMALITY

In this section we consider some properties of the four tests under the basic assumption that all samples are independently drawn from normal distributions. A single class of alternative hypotheses is chosen for study in which (an unknown) one of the populations has greater dispersion than the others which all have identical dispersion. Since computations based on the power function are somewhat intractable for Cochran-type tests, two new functions closely related to the power function are introduced and some numerical values of these functions are computed.

Other information on the power function of tests for homogeneity of variance has recently been given by Pearson (1966) and Leslie and Brown (1966). Their results are concerned with Bartlett's test and the S_{max} / S_{min} and R_{max} / R_{min} tests due to Hartley (1950) and Cadwell (1953), respectively.

Although no accuracy claim is made for the results given in this section, it appears that the four tests are ranked with respect to relative power in the same order as the ranking with respect to efficiency of the estimators upon which they are based. That is, they appear to be ranked in the following order (equivalent test statistics are given in parenthesis):

(1) Cochran's test (S_{max}/S_p), (2) Eisenhart's test ($S_{max}/\bar{S}$),
(3) Mean deviation test ($M_{max}/\bar{M}$), and (4) B-C-T test ($R_{max}/\bar{R}$).

An Alternative Hypothesis

Cochran-type tests of

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$

would seem to be at their "best" against alternatives of the form

$$H_\theta: \sigma_i^2 = \theta^2\sigma^2 , \sigma_j^2 = \sigma^2, \quad \text{for} \quad j = 1,2,\ldots,i-1,i+1,\ldots,k$$

where i, θ and σ are unknown and $\theta > 1$. This class of alternatives is favorable to Cochran-type tests since in this situation they should compare favorably with other types of tests for homogeneity of variance. In fact Cochran's test is the likelihood ratio test for H_0 against alternatives in the class H_θ when the S_i^2 are independently distributed as $\sigma_i^2\chi^2/\nu$. Before stating this formally as a theorem we note the following lemma which is useful in proving the theorem.

LEMMA: Given $a > b > 0$, $c > 0$, $k > 1$, and $b > \frac{a+c}{k-1}$; then

$$R = \frac{b(a+c)^{k-1}}{a(b+c)^{k-1}} > 1 .$$

PROOF: The proof is immediate on noting that $R = 1$ in the limiting case $a = b$ and that

$$\frac{\partial R}{\partial a} = \frac{b(a+c)^{k-2}}{(b+c)^{k-1}a^2}\left[a(k-1) - (a+c)\right] > 0$$

for all $a > b$.

<u>THEOREM</u>: Cochran's test, which is an upper tailed test based on $Y_{max} = S^2_{max} / \sum_{j=1}^{k} S_j^2$, is the likelihood ratio test of H_0 against alternatives of the form H_θ if the S_j^2's are independently distributed as $\sigma_j^2 \chi^2 / \nu$.

PROOF: Let $H_{\theta \ell}$ denote the special case of H_θ in which $\sigma_\ell^2 = \theta^2 \sigma^2$. The proof consists of four steps in which we show:

(1) $\max_{\sigma} L(H_0) = L^* = K \left(\sum_{j=1}^{k} S_j^2 \right)^{-\frac{1}{2}k\nu} \cdot \prod_{j=1}^{k} S_j^{\nu-2}$

(2) $\max_{\theta,\sigma} L(H_{\theta\ell}) = L_\ell^*$, where

$$L_\ell^* = K\, S_\ell^{-\nu} \left(\Sigma_\ell^* S_j^2 \right)^{-\frac{1}{2}(k-1)\nu} \cdot \prod_{j=1}^{k} S_j^{\nu-2} \quad \text{if} \quad S_\ell^2 \geq \Sigma_\ell^* S_j^2 / (k-1)$$

$$= L^* \quad \text{otherwise}$$

(3) $\max_{\ell} L_\ell^* = K\, S_{max}^{-\nu} \left(\Sigma_{max}^* S_j^2 \right)^{-\frac{1}{2}\nu(k-1)} \cdot \prod_{j=1}^{k} S_j^{\nu-2}$

and

(4) the likelihood ratio λ is a monotonic function of Y_{max}.

The notation Σ_ℓ^* denotes summation over all indices except ℓ and Σ_{max}^* denotes summation over all indices except the one corresponding to the largest S^2.

First we state for reference the unrestricted likelihood function.

$$L = K \prod_{j=1}^{k} \sigma_j^{-\nu} S_j^{\nu-2} e^{-\frac{\nu S_j^2}{2\sigma_j^2}} .$$

Step 1. The maximum likelihood estimate of σ^2 under H_0 is readily found to be

$$\hat{\sigma}^2 = \sum_{j=1}^{k} S_j^2 / k$$

and the likelihood function evaluated at its maximum is therefore L^* as defined above.

Step 2. For any given ℓ the maximum likelihood estimates of σ^2 and θ^2 are $\hat{\sigma}^2 = \Sigma_\ell^* S_j^2 / (k-1)$ and $\hat{\theta}^2 = S_\ell^2 / \hat{\sigma}^2$ when θ^2 is <u>not</u> restricted to be greater than or equal to 1. Thus for values of ℓ such that

(7.1.1) $\qquad S_\ell^2 \geq \Sigma_\ell^* S_j^2 / (k-1)$

the value of the likelihood at its maximum is

$$K S_\ell^{-1} \left(\Sigma_\ell^* S_j^2 \right)^{-\frac{1}{2}(k-1)\nu} \prod_{j=1}^{k} S_j^{\nu-2}$$

However, for values of ℓ such that $S_\ell^2 < \Sigma_\ell^* S_j^2 / (k-1)$ the likelihood is maximized under the constraint $\theta^2 \geq 1$ at $\hat{\sigma}^2 = \sum_{j=1}^{k} S_j^2 / k$ and $\hat{\theta}^2 = 1$. This may be seen by showing that the likelihood for <u>given</u> θ^2 is maximized over σ^2 for

$k\hat{\sigma}_\theta^2 = \Sigma_\ell^* S_j^2 + S_\ell^2 / \theta^2$, and then showing that the resulting likelihood evaluated at $\hat{\sigma}_\theta^2$ is a monotonic decreasing function of θ^2 whenever $\theta^2 \geq 1$ and $S_\ell^2 < \Sigma_\ell^* S_j^2 / (k-1)$. Thus the maximum is assumed when $\theta^2 = 1$ which is identical with the case under H_0 for any ℓ.

Step 3. Among the values of ℓ such that inequality (7.1.1) is satisfied, the largest L_ℓ^* may be shown to be the one corresponding to the largest S^2 . This may readily be accomplished using the above lemma and comparing any two L^*'s such that $S_\ell^2 > S_m^2$ and showing that

$$\frac{L_\ell^*}{L_m^*} = \frac{S_\ell^{-\nu} \left(\Sigma_\ell^* S_j^2 \right)^{-\frac{1}{2}(k-1)\nu}}{S_m^{-\nu} \left(\Sigma_m^* S_j^2 \right)^{-\frac{1}{2}(k-1)\nu}} > 1 .$$

The likelihoods for all values of ℓ such that (7.1.1) is not satisfied are identical since $\hat{\theta}^2 = 1$ in those cases. In fact all of the k likelihoods are identical at $\theta^2 = 1$. Now it is readily seen that the likelihood corresponding to the largest S^2 (indeed any S^2 such that 7.1.1 is satisfied) must be at least as large as those with $\hat{\theta}^2 = 1$ since it is maximized for a value of θ other than 1, and at $\theta = 1$ is identical with the others.

Thus $\hat{\sigma}^2 = \Sigma^*_{max} S_j^2 / (k-1)$ and $\hat{\theta}^2 = S^2_{max} / \hat{\sigma}^2$, and the value of the likelihood at its maximum is $\max_{\ell} L^*_{\ell}$ as defined above.

Step 4.

$$\lambda = \frac{\left(\sum_{j=1}^{k} S_j^2\right)^{-\frac{1}{2}k\nu}}{S_{max}^{-\nu}\left(\Sigma^*_{max} S_j^2\right)^{-\frac{1}{2}\nu(k-1)}}$$

$$\lambda^{2/\nu} = \left[\frac{S^2_{max}}{\sum_{j=1}^{k} S_j^2}\right]\left[\frac{\Sigma^*_{max} S_j^2}{\sum_{j=1}^{k} S_j^2}\right]^{k-1}$$

$$= Y_{max}\left[1 - Y_{max}\right]^{k-1}$$

which is monotonic in Y_{max} since $Y_{max} \geq 1/k$ always. This completes the proof.

It has already been noted in Section 5.1 that Truax (1953) has shown that Cochran's test is optimum in a decision theoretic sense where selection of the population with the largest variance is important.

Rejection for the Correct Reason

An important aspect of this particular class of alternative hypotheses is the relative ease with which computations of two functions closely related to the power function may be made. Let

$$P\left(Y_{max} \geq y_\alpha \mid \theta \right)$$

denote the power function of a Cochran-type test of size α for a given alternative H_θ . Then for the class of alternatives considered here it is convenient to break the power function down into two components:

$$P(RC) = P(Y_i \geq y_\alpha \mid \theta)$$

and

$$P(NC) = P\left(Y_{max} \geq y_\alpha \text{ and } Y_i < y_\alpha \mid \theta \right),$$

where the i'th population is the one with the inflated variance. Then $P(RC)$ is the probability that H_θ is rejectable for the "correct" reason and $P(NC)$ is the probability that H_θ is rejectable but not for the correct reason. Note that we say "rejectable" not "rejection," since the occurrence of the event $RC = \{ Y_i \geq y_\alpha \}$ does not necessarily imply $\left\{ Y_i = Y_{max} \right\}$. Others (Truax 1953), and Doornbos and Prins 1956) have described what could be termed rejection for the correct reason. They have considered the "probability of making a correct decision" where the k+1 possible decisions are: to accept H_0 , or to indicate which of the populations has the inflated variance.

The two components RC and NC are mutually exclusive and exhaustive and for reasonably distant alternatives P(RC) constitutes the bulk of the power, as the following theorem indicates.

THEOREM: Let y_α be the critical value for a Cochran-type test of size α. Then for P(RC) defined as above, the power of the test under H_θ bounded by

$$P(RC) < P\left(Y_{max} \geq y_\alpha \mid \theta \right) \leq P(RC) + \alpha(k-1)/k \ .$$

PROOF: The left-hand inequality holds trivially since the event that constitutes the power properly contains the event RC.

The right-hand inequality may be deduced by considering the events

$$A_i = \left\{ Y_i \geq y_\alpha \text{ and } V_i = V_{max} \right\} .$$

(Here and elsewhere in the proof we use the notation V_i, U_i, Y_i, B, and Σ_i^* as described in Section 5.2.) There are k such events A_i which under H_0 are equally likely, each having probability α/k . We now assume for definiteness that under H_θ , the first population has the larger variance. (This is, of course, unknown to the experimenter and is merely a notational convenience.) Then the event RC has the representation

$$RC = A_1 + \left\{ Y_1 \geq y_\alpha \text{ and } V_1 < V_{max} \right\}$$

so

$$P(RC|H_0) \geq \alpha/k \ .$$

Hence

$$P(NC|H_0) \leq \alpha(k-1)/k \ .$$

We now show that $P(NC)$ is maximized under H_0 . First note that

$$NC = \left\{ Y_{max} \geq y_\alpha \text{ and } Y_1 < y_\alpha \right\}$$

$$= \left\{ \frac{U_{max}}{\Sigma_1^* U_j + BU_1} \geq y_\alpha \text{ and } \frac{BU_1}{\Sigma_1^* U_j + BU_1} < y_\alpha \right\} .$$

Hence

$$NC = \left\{ BU_1 \leq \left(U_{max}/y_\alpha - \Sigma_1^* U_j \right) \text{ and } BU_1 < \Sigma_1^* U_j/(y_\alpha^{-1} - 1) \right\}$$

which shows that $P(NC)$ is a monotonic decreasing function of B, completing the proof.

The probability of rejection for the correct reason defined by

$$P\left(Y_{max} \geq y_\alpha \text{ and } V_1 = V_{max} \mid \theta \right),$$

assuming again that the first population has the inflated variance, is slightly smaller than $P(RC)$ but the numerical difference between the two appears to be slight. The new concept is introduced only because computations are simplified considerably. The difference between the two decreases with θ since it is equal to the probability of the event $\left\{ y_\alpha \leq Y_1 < Y_{max} \right\}$ which is contained in the event $\left\{ BU_1 < U_{max} \right\}$, and the probability of the latter event must decrease with θ .

Results on P(RC)

Table 7.1.1 gives the result of some computations of the P(RC) for the four tests under consideration. Also given are approximate values for the power function of Bartlett's test. The significance level .05 is fixed throughout all comparisons. The alternatives were selected to give P(RC) equal to .05, .10, .50, .90 and .95 for Cochran's test.

The approximate values of the P(RC) for the four Cochran-type tests were computed using the methods described in Chapters 5 and 6 for the computation of approximate percentage points. In particular the values for Eisenhart's test were computed using the scaled chi approximations described in Section 6.1 and the values for the mean deviation test were computed using the approximation described in Section 6.2. The values for Cochran's test were computed in the manner outlined by Cochran (1941) and Eisenhart and Solomon (1947) except that α^o was approximated using (5.3.4) instead of (5.3.1a). For the Bliss, Cochran, Tukey (1956) test we used method (A) as outlined in their paper. In all cases we have used approximate nominal critical values computed using the same method of approximation as used here, except possibly for the BCT test where it appears that for $k = 10$ their tabled value is based on a combination of their methods (A) and (B). The values for the power of Bartlett's test were computed using Wilks approximation as outlined by Pearson (1966).

Table 7.1.1

Values of P(RC). The values of θ^2 were chosen to given indicated P(RC) for Cochran's test. The entries in each block are respectively θ^2, P(RC) for the Eisenhart, mean deviation, and Bliss-Cochran-Tukey tests, and the power of Bartlett's test.

P(RC) for Cochran's test	n = 3			5			10		
	k= 3	5	10	k= 3	5	10	k= 3	5	10
	1.94	1.94	2.03	1.53	1.59	1.69	1.31	1.37	1.45
	.049	.049	.049	.050	.050	.050	.050	.050	.048
.05	.050	.048	.046	.048	.047	.045	.047	.046	.044
	.049	.049	.047	.048	.049	.043	.047	.046	.034
	.098	.087	.074	.076	.073	.069	.069	.070	.071
	3.12	2.78	2.74	2.09	2.05	2.11	1.60	1.62	1.70
	.096	.097	.098	.099	.099	.099	.099	.099	.097
.10	.099	.096	.092	.095	.093	.089	.092	.090	.086
	.096	.096	.094	.094	.096	.088	.089	.088	.070
	.129	.118	.101	.112	.104	.096	.105	.102	.098
	16.3	11.4	10.0	6.41	5.45	5.20	3.34	3.12	3.10
	.487	.490	.494	.494	.496	.497	.496	.497	.494
.50	.493	.484	.479	.471	.468	.461	.458	.454	.446
	.484	.487	.486	.473	.479	.465	.436	.438	.398
	.338	.368	.378	.415	.395	.376	.432	.404	.375
	125.	81.0	68.0	23.2	18.4	16.9	7.44	6.59	6.37
	.895	.896	.898	.897	.898	.899	.898	.899	.898
.90	.901	.899	.898	.884	.884	.883	.871	.871	.868
	.893	.894	.895	.883	.886	.882	.849	.851	.833
	.634	.742	.782	.822	.818	.805	.863	.838	.806
	260.	168.	140.	35.4	27.9	25.5	9.52	8.36	8.02
	.948	.948	.949	.949	.949	.949	.949	.949	.949
.95	.952	.951	.951	.941	.941	.941	.932	.932	.930
	.946	.947	.947	.940	.941	.939	.915	.917	.905
	.713	.828	.868	.897	.897	.888	.931	.914	.890

No error checks have been made for the entries in Table 7.1.1 A comparison of the entries for the mean deviation test and the BCT test for $n = 3$ indicates discrepancies of about 1 unit in the second significant figure except for two cases where the differences are 2 and 3 respectively. These entries should theoretically agree by virtue of the theorem in Section 4.2.

The entries for $k = 10$ for the BCT test are somewhat less reliable than the others since a different method of approximation was used in obtaining those critical values.

Discussion

The alternative (i.e., value of θ) needed to produce a given P(RC) for Cochran's test decreases with k for moderately large P(RC), probably for any P(RC) greater than about .20 or .30; and always decreases with n.

The P(RC) for Eisenhart's test is not appreciably lower overall than that of Cochran's test. As k increases from 3 to 5, P(RC) for Eisenhart's test increases for all 3 values of n but as k increases from 5 to 10 the P(RC) increases for $n = 3$ but decreases for $n = 5$ and 10. These changes are very slight and may indeed be due to error in the approximation. The relative efficiency of $\bar{S}$ with respect to S_p decreases with k (see Table 4.2.1) much faster than does the P(RC) of Eisenhart's test relative to Cochran's test.

The difference between the P(RC) for the mean deviation test and that of Cochran's test increases with n, and for $n = 10$ the P(RC) for the mean deviation test is as much as 15% less than that for Cochran's test for $P(RC) = .05$, 11% for $P(RC) = .50$, and 2% for $P(RC) = .95$.

The P(RC) for the BCT test decreases slightly more rapidly with n than does that of the mean deviation test and for $n = 10$ appears to be roughly 1% or so below that for the mean deviation test. (The $n = 10$, $k = 10$ values for the BCT test given here are apparently quite unreliable.)

The values for the power function for Bartlett's test are appreciably below those for the Cochran-type tests for $n = 3$ and higher values of P(RC) . For small values of P(RC), the power function of any test is considerably larger than its P(RC) so comparisons between the power of Bartlett's test and the P(RC) of the Cochran-type tests are not justified for these values. For $n = 5$ and 10 the values for the power function of Bartlett's test are more nearly the same as the P(RC) for the Cochran-type tests with Bartlett's test apparently being more powerful for distant alternatives than the BCT test for $n = 10$ and $k = 3$.

Modified Median Significance Level

Another property closely related to the power function may be called the "median significance level" (MSL). The MSL has been introduced and discussed in some detail in Appendix B and may briefly be defined here as the median of the distribution of the observed significance level of a test for a given alternative.

It has been shown in Appendix B that for one-sided tests the MSL may be computed from the median of the distribution of the test statistic under the given alternative. Thus in the present notation, the MSL of a CTT for an alternative H_θ could be computed as the solution p of the implicit equation

$$P\left(Y_{max} \geq y_p \mid \theta \right) = \tfrac{1}{2} .$$

It has also been shown in Appendix B that the MSL may be considered the significance level for which there is power equal to $\frac{1}{2}$ for a given alternative. In light of this relation and since it was shown in the preceding theorem that for moderately distant alternatives the P(RC) constitutes the bulk of the power, the relative simplicity of P(RC) computations leads us to suggest a modified version of the MSL based on P(RC) . That is, let p be the solution of the implicit equation

$$P\left(Y_1 \geq y_p \mid \theta \right) = \tfrac{1}{2}$$

where we again assume for notational convenience that the first population is the one with the inflated variance. Then the solution p may be termed the "modified median significance level" (MMSL).

The MMSL may be interpreted as an approximation to the MSL, or as an answer to the question "at what significance level is there a 50-50 chance of having H_0 rejectable for the correct reason?" Small values of the MMSL are, of course, more desirable.

Table 7.1.2 gives some computed values of the MMSL which were obtained by using the same approximations described earlier in this section. The alternatives were chosen to give MMSL values of .25, .05, .01, .001 and .0001 for Cochran's test. No error checks have been made for Table 7.1.2 but the results seem somewhat erratic suggesting that comparisons should be made with caution. For $n = 3$ the entries for the mean deviation test and the BCT test should be identical but discrepancies as large as 6 units in the second significant figure may be noted. The larger error in these results as compared with the error in Table 7.1.1 is probably due to the fact that MMSL computations depend more sensitively on the tail of the approximating distribution.

Table 7.1.2

Values of the "modified median significance level." Alternatives (i.e. values of θ^2) were chosen to give specified MMSL for Cochran's test. The upper entry in each block is the value of θ^2 and the three lower entries are for the Eisenhart, mean deviation, and Bliss-Cochran-Tukey tests.

MMSL for Cochran's test	n = 3			5			10		
	k=3	5	10	k=3	5	10	k=3	5	10
.25	5.95	5.89	6.24	3.35	3.39	3.66	2.20	2.26	2.42
	.257	.251	.260	.246	.255	.255	.244	.252	.252
	.257	.257	.279	.264	.284	.303	.274	.297	.319
	.260	.254	.265	.261	.275	.283	.295	.322	.350
.05	16.3	11.4	9.99	6.41	5.45	5.20	3.34	3.12	3.10
	.054	.055	.054	.052	.052	.052	.050	.051	.051
	.051	.057	.062	.059	.064	.072	.065	.070	.078
	.055	.056	.056	.059	.060	.062	.076	.082	.090
.01	39.4	19.7	14.4	10.9	7.87	6.82	4.65	3.99	3.77
	.011	.012	.012	.011	.011	.011	.010	.010	.010
	.010	.012	.014	.013	.015	.017	.015	.017	.019
	.012	.012	.012	.013	.013	.014	.019	.021	.023
.001 ($X10^1$)	130.	39.2	22.3	21.6	12.3	9.37	6.93	5.35	4.73
	.012	.013	.013	.011	.012	.011	.010	.011	.010
	.010	.013	.016	.014	.017	.022	.018	.021	.026
	.013	.014	.014	.015	.016	.016	.027	.030	.033
.0001 ($X10^2$)	416.	73.8	32.4	40.4	18.1	12.2	9.84	6.85	5.72
	.013	.015	.014	.012	.012	.012	.011	.011	.011
	.009	.013	.019	.014	.020	.028	.021	.027	.035
	.014	.016	.016	.017	.018	.019	.039	.042	.046

Discussion

The alternative (i.e., value of θ) needed to produce an MMSL of .25 is roughly constant for fixed n (in the range 3 to 10) as k increases from 3 to 10; and is a decreasing function of n over the range of the table. The alternative needed to produce smaller values of the MMSL decreases both with n and with k.

There do not appear to be any appreciable differences among the four tests for $n = 3$ and 5 but for $n = 10$ the BCT and mean deviation tests seem to fall sufficiently far behind Cochran's and Eisenhart's tests to possibly make a practical difference. The MMSL and P(RC) results by and large seem to provide pretty much the same impressions as to the relative behavior of the four tests.

To sum up briefly it seems that for the small sample sizes considered here there is little reason to choose one or the other of the four tests on the basis of statistical efficiency.

7.2 PROPERTIES UNDER NON-NORMALITY

In this section we give some results concerning the properties of the four Cochran-type tests and Bartlett's test when all samples are from identical but non-normal distributions. Most of these results are based on a modified degrees of freedom approximation originally suggested by Le Roux (1931). This approximation has been used by Box (1953a) in assessing the non-robustness of Bartlett's (1937) test and Hartley's (1950) test. In addition the results of a limited sampling study are given for the above tests and several others, for samples from four non-normal distributions.

Most results on the effect of non-normality on tests for homogeneity of variance have been obtained by Box and co-workers. See for example, Box (1953a, 1953b), Box and Andersen (1955), and Box and Tiao (1962, 1964a, 1964b). The asymptotic effect of non-normality on a general class of tests has also been considered by Laue (1965). This class of tests includes Stevens' (1936) test, Bartlett's (1937) test, Cochran's (1941) test, Neyman's (1941) test, Hartley's (1950 test, Bartlett and Kendall's (1946) test, and Eisenhart's test. Laue shows that all tests in this class are asymptotically (in n) equally effected by non-normality.

Our approach differs in that we have considered only small samples and have attempted to apply "modified degrees of freedom" approximations to the tests based on the mean deviation and range. Our conclusions are basically the same as Box's and Laue's in that all of the tests appear to be quite sensitive to non-normality. However, there do appear to be some differences, at least in small samples, in the extent to which the various tests are affected by different types of non-normality. These results are summarized below.

Concept and Importance of Robustness

Before beginning a detailed discussion of the properties of Cochran-type tests for homogeneity of variance, it seems important to digress for a moment and discuss briefly the basic concept of robustness and its relevance in situations where one might consider testing the equality of variances.

The term "robust" was first applied to the properties of statistical tests by Box (1953a), but the basic concept was not new since Pearson and Adyanthaya (1928) and others had investigated the properties of certain tests when the underlying assumptions were not fully satisfied. Fridshal and Posten (1966) have recently given an extensive bibliography of the literature on robustness, and Hatch and Posten (1966) have given a comprehensive survey of results on the robustness of the Student procedure for testing means. In their survey many concepts are discussed which are pertinent in a more general discussion of the robustness problem.

A test is said to be "robust" with respect to a particular type of departure from assumptions if the probability that the null hypothesis is rejected does not vary "too much" under "moderate" departures from the assumptions. The "too much" and "moderate" have not yet been quantitively defined. In this work we are only concerned with the robustness of the Cochran-type tests when the null hypothesis is true and when all assumptions except the assumption of normality are satisfied.

With respect to tests for homogeneity of variance, it has been argued that robustness with respect to non-normality is not necessarily a desirable property. For example, in an industrial or calibration situation where one is striving to bring a process under control the argument has been advanced (by Ellis R. Ott, and others in personal communication) that departures from normality frequently indicate that the process is not fully debugged, and that if one looks for the source of trouble, one is usually rewarded. This feeling seems to be prevalent in the quality control literature. The legendary success of statistical quality control which has been achieved partially through the use of S and R charts (which are usually considered to be basically tests for homogeneity of variance) can be interpreted as a strong practical argument <u>against</u> the desirability of robustness with respect to non-normality in this context.

However, from a theoretical point of view it seems important to distinguish formally (a) underlying assumptions that one does not wish to test, and (b) the full null hypothesis which is to be tested. In the analysis of data from well-controlled but non-normal distributions, it may sometimes be preferable to use a robust test instead of trying to find a normalizing transformation.

It may be remarked that in the ideal case of perfect robustness to a subset of assumptions, that is, complete independence from these assumptions, the so called assumptions cease to be assumptions and become irrelevant considerations. One would then be in the fortunate situation of having a "similar region" test with respect to that subset of "assumptions."

Basic Method of Approximation

Here as in most cases, assessing the effect of non-normality on the properties of statistical tests is very difficult and one must resort to fairly rough approximations. The basic approximation used here stems from the observation of Le Roux (1931) that the distribution of S^2 in samples from certain non-normal distributions is very nearly like that of S^2 in samples of a modified size from a normal distribution.

In this section we use a slightly "modified-generalized" version of his approximation. We assume that the distribution of an estimator T in samples of size n from a non-normal distribution is the same, to a first order of approximation, as that of T in samples of size n^* from a normal distribution, where n^* is such that $CV^2(T)$ is given correctly. Since we assume throughout that the component estimators in any given test all have the same distribution, the expectation of T is irrelevant to the properties of the tests and the first integral moment of T that can influence the properties of the test is equivalent to $CV^2(T)$.

When the component estimators are sample variances (Cochran's test) the value of n^* is a simple function of β_2 and is given exactly by

$$n^* - 1 = (n-1)\left[1 + \frac{1}{2}\,\frac{n-1}{n}\,(\beta_2 - 3)\right]^{-1} .$$

We have also given some evidence in Section 4.5 that this same modification of the sample size also seems to provide a fairly good approximation to the coefficient of variation of S for samples from a number of discrete distributions.

In Tables 4.7.1 and 4.7.2 we have given three different modifications of the sample size for R, depending on whether the observations came from TRV-small, TRV-large, or Cox's "average" distributions. Different conclusions may be drawn for samples from the different classes of distributions.

No relationship between β_2 and $CV^2(M)$ is available at this time, although we have seen in Section 4.5 that the $CV^2(M)$ behaves much like $CV^2(S)$ in small samples from non-normal distributions.

Results Based on the Approximation

Table 7.2.1 gives approximate values for the true significance levels of the four tests corresponding to a nominal significance level of .05. Values are given for n and k equal to 3, 5, and 10 for several values of n^*. These approximate values were computed using the same approximations as were used in the previous section.

The values for Bartlett's test were computed using Bartlett's (1937) approximation based on a scaled chi-square with k-1 degrees of freedom. Let L and C_n be defined by

$$L = (n-1)\left[k\,\ell n\left(\sum_{j=1}^{k} S_j^2 / k \right) - \ell n\left(\prod_{j=1}^{k} S_j^2 \right) \right]$$

and

$$C_n = 1 + [\ k+1\ ] / [\ 3k(n-1)\ ] \ .$$

Under normality L/C_n is approximately distributed as χ^2_{k-1}, but in a non-normal situation where the S^2's behave as if they were based on a sample size n^*, L/C_n is approximately distributed as

$$\frac{(n-1)}{(n^*-1)} \frac{C_{n^*}}{C_n} \cdot \chi^2_{k-1} \ .$$

No accuracy claim is made for the results given in Table 7.2.1, but it seems certain that the error in the computed values for a given n^* is minor compared to the error involved in extrapolating these results to a population described essentially only by a given value of β_2. Spot checks using the results of a sampling study based on the approximation indicate that the

Table 7.2.1

Values for the "modified degrees of freedom" approximation to the true significance level corresponding to a nominal significance level of .05. Values in a block are for the Cochran, Eisenhart, mean deviation, BCT, and Bartlett tests, respectively. β_2 values for the Le Roux approximation are shown in parenthesis.

n = 3

n^*	k= 3	5	10
2	.20	.21	.24
(6)	.22	.26	.35
	.22	.26	.35
	.22	.26	.34*
	.30	.42	.62
4	.013	.013	.011
(2)	.012	.010	.007
	.013	.011	.008
	.013	.011	.007*
	.010	.006	.002

n = 5

n^*	k= 3	5	10
2	.41	.47	.57
(10)	.47	.57	.80
	.47	.57	.80
	.45	.55	.75*
	.56	.76	.95
3	.19	.22	.26
(5.5)	.21	.26	.35
	.20	.23	.31
	.20	.24	.30*
	.26	.36	.56
4	.097	.103	.113
(3.8)	.102	.113	.133
	.097	.107	.121
	.097	.108	.111*
	.12	.14	.20
8	.0075	.0062	.0045
(1.9)	.0065	.0047	.0029
	.0073	.0055	.0034
	.0087	.0069	.0035*
	.0043	.0018	.0004

n = 10

n^*	k= 3	5	10
3	.44	.50	.68
(11)	.46	.61	.85
	.43	.57	.79
	.42	.55	.72*
	.56	.77	.95
4	.31	.36	.48
(7.4)	.33	.43	.60
	.31	.40	.56
	.29	.37	.45*
	.40	.57	.81
5	.22	.26	.33
(5.8)	.24	.30	.40
	.22	.28	.37
	.21	.25	.28*
	.28	.41	.62
7	.121	.132	.155
(4.1)	.124	.144	.172
	.121	.138	.163
	.111	.126	.113*
	.14	.19	.27
17	.0071	.0056	.0038
(2)	.0064	.0046	.0026
	.0069	.0050	.0030
	.0120	.0096	.0037*
	.0044	.0017	.0004

* Critical values based on a different method of approximation.

reported values for the Bartlett, Cochran, and Eisenhart tests are accurate at least within two units in the second significant figure.

Interpretation of Results Based on the Approximation

For any given value of n^* there seems to be virtually no difference in the results for the four Cochran-type tests for $k = 3$, while for $k = 10$ and the smaller values of n^* there is a slight difference with Cochran's test appearing to be most stable. However, all four of the tests deviate considerably from the nominal .05 value for values of n^* different from n, suggesting severe overall non-robustness.

Supposing for a moment that the coefficients of variation of S^2, S, M and R were always equally affected by non-normality, these first order results would seem to indicate that there was little difference in the robustness of the tests, but that if a choice had to be made Cochran's test based on S^2 would seem to have the edge. However, we have shown in Sections 4.5 and 4.6 that the coefficients of variation are not all equally affected by non-normality and the interpretation of these results based on the modified degrees of freedom approximations is not nearly so clear. For example, in samples from TRV-small distributions these results indicate that R should be slightly better than S^2 since we see from Table 4.7.2 that for n and $k = 5$ and larger β_2 (say $\beta_2 = 9$),

n^* would be 3 for R but about 2 for S^2 giving significance levels of roughly .25 and .50, respectively. For TRV-large distributions the situation would be reversed since for n and $k = 5$ and $\beta_2 = 8$, n^* would be 2 for R and about 2.5 for S^2 giving significance levels of roughly .55 and .40 respectively.

Thus we have two related families of distributions, TRV-small and TRV-large, for which the Bliss-Cochran-Tukey test based on R appears to have substantially different robustness properties. The TRV-small family has been shown in Section 4.6 to have a CV(R) much like the Burr and miscellaneous distributions while the TRV-large family has a CV(R) that is quite different, being more like that of the discrete distributions. It seems that to the validity of the approximation, the Bliss-Cochran-Tukey test should be somewhat more robust than Cochran's test for small samples from most "reasonable" continuous distributions (i.e., those similar to the TRV-small distributions) while for samples from distributions where discreteness is a real problem (say less than about 8 distinct values) Cochran's test would be preferred.

The approximation suggests that all of the tests become increasingly non-robust as the number of sample increases. It appears from these results that Cochran's test always becomes increasingly non-robust as the sample size increases, and that the BCT test becomes increasingly non-robust with n for TRV-small samples but becomes slightly more robust with n for TRV-large samples.

Box (1953a) has given approximate values, computed in an analogous manner, for the true significance level of Bartlett's test for $\beta_2 = 4$ and a nominal significance level of .05 with $n = 12$. From Table 3.1, $n^* = 13.55$. The values given by the approximation are .175, .252, and .373 for $k = 5$, 10 and 20 respectively. The corresponding results for Cochran's test are .130, .155, and .185. (These differences were not borne out by the sampling study.)

Results of the Sampling Study

A limited sampling study was undertaken in order to assess the validity of the modified degrees of freedom approximations. The results of this study are reported in Tables 7.2.2 and 7.2.3. Results have been obtained for samples from chi-square distributions with 2, 4 and 12 degrees of freedom and from the TRV-small distribution with $\lambda = .0802$ (see Section 4.4 for more details on the TRV distributions). Table 7.2.4 lists the critical values used in obtaining the various results.

The chi-square distribution with 12 degrees of freedom was chosen as being a "reasonable" and convenient distribution having a β_2 value of 4, the value used by Box (1953a) in his illustration of the non-robustness of Bartlett's test. Samples of size $n = 21$ were drawn from this distribution for $k = 5$, 10 and 20. We then counted the number of sets in which H_0 was

Table 7.2.2

Approximate probabilities of rejecting H_0 at the .05 level for samples of size n=21 from parent distributions having $\beta_2=4$. Upper entry gives the "modified degrees of freedom" approximation and lower entry gives result of sampling study based on samples from the χ^2_{12} distribution.

k	Bartlett	Cochran	Eisenhart	Hartley	No. of Samples	2σ limits
5	.175 .15	.130 .14	.13	.14	1200	.02
10	.252 .20	.155 .19	.18	.19	400	.04
20	.373 .31*	.185 .28	.23		900	.03

* 400 samples; $2\sigma = .05$

Table 7.2.3

Empirical sampling estimates of the true significance level corresponding to a nominal significance level of .05.

Parent Distr.	β_2	n	k	No. of samples	a	b	c	d	e	f	g	h
χ^2_2	9	5	3	1000	.26	.25	.23	.25				
		5	5	1000	.36	.29	.31	.32				
		5	10	1000	.55	.42	.39	.47				
		10	3	1000	.31	.26	.31	.27				
		10	5	2000	.48	.38	.46	.42				
		10	10	1000	.71	.51	.59	.55				
χ^2_4	6	5	3	1000	.14	.15	.12	.14	.14	.12	.11	.13
		5	5	1000	.18	.17	.14	.18	.17	.14	.15	.16
		5	10	2000	.29	.27	.18	.30	.24	.16	.16	.25
		10	3	1000	.18	.17	.17	.17	.15	.15	.14	.13
		10	5	1000	.28	.25	.25	.26	.23	.22	.18	.19
		10	10	1000	.41	.33	.31	.35	.32	.29	.23	.24
TRV $\lambda = .0802$	6	5	5	1000	.17	.18	.13	.19	.19	.14	.11	.15
		10	3	1000	.17	.17	.16	.18	.18	.18	.12	.13
		10	10	1000	.36	.30	.27	.31	.30	.30	.16	.19

(a) Bartlett

(b) Cochran

(c) Hartley

(d) Eisenhart

(e) Bliss-Cochran-Tukey

(f) R-Max

(g) M-Max

(h) Mean deviation test

Table 7.2.4

Summary of approximate critical values used in Chapter 7.

		n = 5			n = 10			n = 21		
		k=3	5	10	k=3	5	10	k=5	10	20
(a)	BARTLETT	6.61	10.37	18.38	6.28	9.90	17.60	9.64	17.23	30.67
(b)	COCHRAN	.7457	.5441	.3308	.6167	.4239	.2437	.3457	.1908	.1031
(c)	HARTLEY	15.5	25.2	44.6	5.34	7.11	9.91	3.54	4.37	--
(d)	EISENHART	.557	.365	.199	.477	.305	.163	.270*	.142*	.074*
(e)	BLISS-C-T	.563	.369	.204	.489	.314	.172			
(f)	R-MAX	4.018	5.149	6.892	2.452	2.855	3.404			
(g)	M-MAX	4.04	5.13	6.85	2.39	2.77	3.29			
(h)	MEAN DEV.	.563	.371	.204	.484	.311	.167			

* n = 20

(a) Pearson and Hartley (1954) except for n=21 which were computed using Bartlett's (1937) approximation.

(b) Computed as described in Chapter 5 using (5.3.4)

(c) Pearson and Hartley (1954)

(d) Table 6.1.3 except Scaled Chi Approximation

(e) Bliss, Cochran, Tukey (1956)

(f) Leslie and Brown (1966)

(g) Cadwell (1953, p.346)

(h) Table 6.2.1

rejected by Bartlett's test, Cochran's test, Eisenhart's test and Hartley's test. These results are given in Table 7.2.2. For Eisenhart's test the critical value corresponding to $n = 20$ rather than 21 was used probably causing a slight distortion of the results. The value for $n = 21$ is smaller by roughly 1 unit in the third decimal place so the observed counts are smaller than they should be. There do not appear to be any significant differences among the results for the four tests in this situation, except for a possible indication of a slight increase in robustness for Eisenhart's test. However, this discrepancy is more likely due to the use of inappropriate critical values for Eisenhart's test.

However, when these sampling results are compared with the results from the modified degrees of freedom (mdf) approximation, serious discrepancies are apparent. The two methods are in rough agreement for $k = 5$ but for $k = 10$ the mdf value for Bartlett's test is .05 larger than the χ^2_{12} result while the mdf value for Cochran's test is .05 smaller than the χ^2_{12} result. For $k = 20$ the discrepancy has increased to .06 for Bartlett's test and .10 for Cochran's test. Thus the difference between the two tests is indicated to be much larger by the mdf approximation than is borne out for the χ^2_{12} distribution. This discrepancy may be due to fundamental differences between the χ^2_{12} distribution and a parent distribution that would give rise to a distribution of S^2's similar in nature to a χ^2 distribution with 13.55 degrees of freedom.

Table 7.2.3 gives the results of the sampling study for n and k equal to 3, 5, and 10. For samples from a chi-square distribution with 2 degrees of freedom (i.e. the exponential distribution) there does not appear to be any clear cut ranking of the tests except that Bartlett's test appears to be slightly less satisfactory overall than the Cochran, Eisenhart and Hartley tests. The χ^2_2 distribution is quite an extreme departure from normality having a β_2 value of 9 and being extremely skew. For the tests based on S^2 the n^* values are found from Table 3.1 to be 2.18 for $n = 5$ and 3.43 for $n = 10$. Rough interpolation in Table 7.2.1 shows that for $n = 5$ the mdf values would be approximately (sampling results are given in parentheses)

	k = 3	5	10
Bartlett	.50 (.26)	.70 (.36)	.85 (.55)
Cochran	.35 (.25)	.40 (.29)	.50 (.42)

and for $n = 10$ would be approximately

	k = 3	5	10
Bartlett	.50 (.31)	.70 (.48)	.90 (.71)
Cochran	.40 (.26)	.45 (.38)	.60 (.51) .

A comparison between these values and the results of the sampling study again shows serious disagreement. Here both the Bartlett and Cochran mdf values are larger than the χ_2^2 values, with the disagreement being greatest for Bartlett's test. However, the mdf and χ_2^2 results do agree as to the ranking of the tests with respect to robustness.

Results for samples from the chi-square and TRV-small distributions both having a β_2 value of 6 are also given in Table 7.2.3. It seems here that for $n = 5$ the three tests of the form $\max V_i / \min V_i$ are less affected by non-normality while for $n = 10$ the two tests based on the mean deviation appear to have a slight edge over the others. The remaining tests appear to be pretty much alike here except that for k and n equal 10, Bartlett's test appears to be slightly worse than the others.

The results for the two different parent distributions seem to agree quite nicely overall. Possible exceptions are Bartlett's test for $n = k = 10$ and the two tests based on the mean deviation. The TRV-small distribution seems to be more favorable toward the mean deviation tests (and possibly all of the tests) for $n = k = 10$.

We have given in Table 4.7.1 the necessary modification of the sample size for the range in samples from the TRV-small distribution with $\beta_2 = 6$. This modification is $n = 3.5$ for $n = 5$ and $n^* = 5$ for $n = 10$. For the tests based on S^2,

Table 3.1 indicates that $n^* = 2.82$ for $n = 5$ and $n^* = 4.83$ for $n = 10$. Thus interpolating roughly in Table 7.2.1 we find that the mdf approximation gives for $n = 5$:

	k = 3	5	10
Bartlett	.32 (.14)	.44 (.18)	.64 (.29)
Cochran	.24 (.15)	.26 (.17)	.32 (.27)
BCT	.15 (.14)	.17 (.17)	-- (.24)

and for $n = 10$:

	k = 3	5	10
Bartlett	.31 (.18)	.44 (.28)	.64 (.38)
Cochran	.24 (.17)	.28 (.25)	.36 (.32)
BCT	.21 (.18)	.25 (.23)	-- (.30) .

In this situation the mdf results are again considerably larger than the sampling results for the two tests based on S^2, particularly for Bartlett's test. The results for Cochran's test tend toward agreement as k increases from 3 to 5 to 10. The agreement between the two types of results for the BCT test is quite good.

Discussion

It appears that for the situations considered here there is little evidence of any appreciable difference in the robustness of the four Cochran-type tests. Insufficient evidence is available at the present time on the properties of the mean deviation to interpret the modified degrees of freedom results for the mean deviation test but the few sampling results that are available do indicate that it might have a slight advantage over the other Cochran-type tests.

Among the whole group of tests considered, Bartlett's test appeared to be the most non-robust overall. The max V_i / min V_i tests appear to be relatively more robust overall than one might intuitively expect. For $n = 5$ they seemed to be uniformly superior for the two distributions considered. Whether or not these results would carry over in any systematic manner to other parent distributions is not readily apparent. It does appear at this time that a knowledge of β_2 alone is not sufficient to give precise results, but by and large a knowledge of the value of β_2 does seem to be sufficient to provide qualitatively correct information on the types of departures one might expect.

8. ACKNOWLEDGEMENTS

I wish to thank all of those at the National Bureau of Standards and Rutgers University who have helped to make this endeavor possible, and in particular Dr. Joan R. Rosenblatt whose assistance and guidance contributed immeasureably to the substance of this work. Special thanks must also go to Dr. Ellis R. Ott for his overall inspirational encouragement throughout my statistical education; to Dr. Churchill Eisenhart who contributed many basic ideas; to Dr. David Hogben for assistance in solving a number of miscellaneous problems, and to Mrs. Ruth N. Varner for assistance in implementing the statistical computer programs used throughout.

I also wish to express my gratitude to Miss Bonnie Connor whose typing skill is manifest throughout this work. Many additional thanks go to my wife, Laurie, who typed numerous preliminary drafts and otherwise encouraged and assisted me in completing this manuscript.

9. REFERENCES

BAHADUR, R. R. (1960). Stochastic comparison of tests. Ann. Math. Statist., 31, 276-295.

BAKER, G. A. (1935). Note on the distributions of the standard deviations and second moments of samples from a Gram-Charlier population. Ann. Math. Statist., 6, 127-130.

BARGMANN, R. E. and GHOSH, S. P. (1963). Statistical distribution programs for a computer language. IBM Reports, RC 1094.

BARTLETT, M. S. (1937). Properties of sufficiency and statistical tests. Proc. Roy. Soc., A, 160, 268-282.

BARTLETT, M. S. and KENDALL, D. G. (1946). The statistical analysis of variance heterogeneity and the logarithmic transformation. J. Roy. Statist. Soc., B, 8, 128-138.

BECHOFER, R. E. (1960). A multiplicative model for analyzing variances which are affected by several factors. J. Amer. Statist. Asso., 55, 245-264.

BENNETT, B. M. (1955). On the joint distribution of the mean and standard deviation. Ann. Inst. Statist. Math., Tokyo, 7, 63-66.

BIRNBAUM, A. and DUDMAN, J. (1963). Logistic order statistics. Ann. Math. Statist., 34, 658-663.

BLAND, R. P., GILBERT, R. D., KAPADIA, C.H. and OWEN, D. B. (1966). On the distributions of the range and mean range for samples from a normal distribution. Biometrika, 53, 245-248.

BLISS, C. I., COCHRAN, W. G. and TUKEY, J. W. (1956). A rejection criterion based upon the range. Biometrika, 43, 418-522.

BOX, G. E. P. (1953a). Non-normality and tests on variances. Biometrika, 40, 318-335.

BOX, G. E. P. (1953b). Regions for tests of kurtosis. Biometrika, 40, 465-468.

BOX, G. E. P. and ANDERSEN, S. L. (1955). Permutation theory in the derivation of robust criteria and the study of departures from assumption. J. Roy. Statist. Soc., B, 17, 1-26.

BOX, G. E. P. and TIAO, G. C. (1962). A further look at robustness via Bayes' theorem. Biometrika, 49, 419-432.

BOX, G. E. P. and TIAO, G. C. (1964a). A Bayesian approach to the importance of assumptions applied to the comparison of variances. Biometrika, 51, 153-168.

BOX, G. E. P. and TIAO, G. C. (1964b). A note on criterion robustness and inference robustness. Biometrika, 51, 169-174.

BURR, I. (1955). Calculation of exact sampling distribution of ranges from discrete population. Ann. Math. Statist., 26, 530-532.

BURR, I. W. (1966). The effect of non-normality on constants for measurement control charts using ranges. Purdue University, Department of Statistics Mimeograph Series, No. 93. Published as Burr (1967).

BURR, I. W. (1967). The effect of non-normality on constants for $\bar{X}$ and R charts. Indust. Qual. Control, 23, 563-569.

CADWELL, J. H. (1953). Approximating to the distributions of measures of dispersion by a power of χ^2. Biometrika, 40, 336-346.

CADWELL, J. H. (1954). The probability integral of range for samples from a symmetrical unimodal population. Ann. Math. Statist., 25, 803-806.

CADWELL, J. H. (1954a). The statistical treatment of mean deviation. Biometrika, 41, 12-18.

CARLSON, P. G. (1958). A recurrence formula for the mean range for odd sample sizes. Skand. Aktuarietidskr., 41, 55-56. Note the typographical error in the lemma on page 55. Change $\sum_{i=1}^{n-1}$ to $\sum_{i=0}^{n-1} \binom{n}{i}$.

CHAND, U. (1949). Formulas for the percentage points of the distribution of the arithmetic mean in random samples from certain symmetrical universes. J. Res. Nat'l. Bur. Stds., 43, 79-80.

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. Ann. Math. Statist., 23, 493-507.

CHURCH, A. E. R. (1926). On the means and squared standard deviations of small samples from any population. Biometrika, 18, 321-394.

COCHRAN, W. G. (1941). The distribution of the largest of a set of estimated variances as a fraction of their total. Ann. Eugenics, London, 11, 47-52.

COX, D. R. (1954). The mean and coefficient of variation of range in small samples from non-normal population. Biometrika, 41, 469-481.

COX, D. R. (1955). Discussion on the paper by Dr. Box and Dr. Anderson. J. Roy. Statist. Soc., B, 17, 28-29.

CRAIG, A. T. (1932). The simultaneous distribution of mean and standard deviation in small samples. Ann. Math. Statist., 3, 126-140.

CRAIG, C. C. (1929). Sampling when the parent population is of Pearson's Type III. Biometrika, 21, 287-293.

DARLING, D. A. (1952). On a test for homogeneity and extreme values. Ann. Math. Statist., 23, 450-456.

DAVID, H. A. (1954). The distribution of the range in certain non-normal populations. Biometrika, 41, 463-468.

DAVID, H. A. (1956). On the application to statistics of an elementary theorem in probability. Biometrika, 43, 85-91.

DAVID, H. A. (1962). Order statistics in short-cut tests. Chapter 7A of Sarhan and Greenberg (1962).

DEMPSTER, A. P. and SCHATZOFF, M. (1965). Expected significance level as a sensitivity index for test statistics. J. Amer. Statist. Asso., 60, 420-436.

DOORNBOS, R. (1956). Significance of the smallest of a set of estimated normal variances. Statistica Neerlandica, 10, 17-26.

DOORNBOS, R. and PRINS, H. J. (1956). Slippage tests for a set of Gamma-variates. Indag. Math., 18, 329-337.

DUNLAP, H. F. (1931). An emperical determination of the distribution of means, standard deviations and correlation coefficients drawn from rectangular populations. Ann. Math. Statist., 2, 66-81.

EISENHART, C. (1947). Effects of rounding or grouping data. Chapter 4 of Eisenhart, Hastay, and Wallis (1947).

EISENHART, C., HASTAY, M. W. and WALLIS, W. A. (1947). Selected Techniques of Statistical Analysis. McGraw-Hill Book Co., New York, New York.

EISENHART, C. and SOLOMON, H. (1947). Significance of the largest of a set of sample estimates of variance. Chapter 15 of Eisenhart, Hastay and Wallis (1947).

FELLER, W. (1957). An Introduction to Probability Theory and Its Applications. 2nd ed., John Wiley and Sons, New York.

FISHER, R. A. (1920). A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error. Monthly Notices of the Royal Astronomical Society, 80, 758-770.

FISHER, R. A. (1929, 1950). Tests of significance in harmonic analysis. (1929) Proc. Roy. Soc., A, 125, 54-59, reproduced with corrections (1950) as paper No. 16 of R. A. Fisher, Contributions to Mathematical Statistics, John Wiley and Sons, New York.

FRIDSHAL, D. and POSTEN, H. O. (1966). Bibliography on statistical robustness and related topics. Research Report No. 16, Dept. of Stat., University of Connecticut.

GEARY, R. C. (1935). The ratio of the mean deviation to the standard deviation as a test of normality. Biometrika, 27, 310-332.

GEARY, R. C. (1966). The average critical value method for adjudging relative efficiency of statistical tests in time series regression analysis. Biometrika, 53, 109-119.

GEPHART, L. S. (1955). A study of the effect of non-normality on sampling distributions of the range. WADC Tech. Rept., 55-128.

GODWIN, H. J. and HARTLEY, H. O. (1945). Tables of the probability integral of the mean deviation in normal samples. Biometrika, 33, 259-265.

GOVINDARAJULU, Z. (1966). Best linear estimates under symmetric censoring of the parameters of a double exponential population. J. Amer. Statist. Asso., 61, 248-258.

GOVINDARAJULU, Z. and EISENSTAT, S. (1965). Best estimates of location and scale parameters of a chi (1 d.f.) distribution, using ordered observations. Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs., 12, No. 4, 149-164.

GOVINDARAJULU, Z. and HUBACKER, N. W. (1964). Percentiles of order statistics in samples from uniform, normal, chi (1 d.f.) and Weibull populations. Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs., 11, 18-44.

GUPTA, S. S. (1960). Order statistics from the gamma distribution. Technometrics, 2, 243-262.

GUPTA, S. S., QUREISHI, A. S., and SHAH, B. K. (1967). Best linear unbiased estimators of the parameters of the logistic distribution using order statistics. Technometrics, 9, 43-56.

GUPTA, S. S. and SHAH, B. (1965). Exact moments and percentage points of the order statistics and the distribution of the range from the logistic distribution. Ann. Math. Statist., 36, 907-920.

HARRIS, T. E. and TUKEY, J. W. (1949). Development of large-sample measure of location and scale which are relatively insensitive to contamination. Statist. Res. Group, Princeton Univ., Memorandum Report 31.

HARTER, H. L. (1960). Tables of range and studentized range. Ann. Math. Statist., 31, 1122-1147.

HARTER, H. L. (1963). Percentage points of the ratio of two ranges and power of the associated test. Biometrika, 50, 187-194.

HARTER, H. L. (1964). Expected values of exponential, Weibull, and gamma order statistics. Aerospace Res. Lab., Tech. Rept. 64-31, AD-436763, 106 pages.

HARTER, H. L. and CLEMM, D. S. (1959). The probability integrals of the range and studentized range, I. Wright Air Development Center, Ohio. Tech. Rept. 58-484, AD-215024, 139 pages.

HARTLEY, H. O. (1950). The maximum F-ratio as a shortcut test for heterogeneity of variance. Biometrika, 37, 308-12.

HARTLEY, H. O. and DAVID, H. A. (1954). Universal bounds for mean range and extreme observation. Ann. Math. Statist., 25, 85-99.

HARTLEY, H. O. and PEARSON, E. S. (1946). Tables for testing the homogeneity of a set of estimated variances. Biometrika, 33, 296-304.

HASTINGS, C. MOSTELLER, F., TUKEY, J. W., and WINSOR, C. P. (1947). Low moments for small samples: A comparative study of order statistics. Ann. Math. Statist., 18, 413-426.

HATCH, L. O. and POSTEN, H. O. (1966). Robustness of the Student-Procedure: A survey. Res. Rept. No. 24, Dept. of Statist., Univ. of Connecticut.

HERREY, E. (1965). Confidence intervals based on the mean absolute deviation of a normal sample. J. Amer. Statist. Asso., 60, 257-269.

HOGBEN, D. (1963). Some properties of Tukey's test for non-additivity. Unpublished Ph.D. Thesis, Rutgers - The State University.

HOGBEN, D., PINKHAM, R. S., and WILK, M. B. (1962). Some properties of Tukey's test for non-additivity. Abstract in Ann. Math. Statist., 33, 1492-1493.

HUME, M. W. (1965). The distribution of statistics expressable as maxima. Virginia J. of Sci., 16, 120-127.

IRICK, P. (1950). A geometric method for finding the distribution of standard deviations when the sampled population is arbitrary (preliminary report). Ann. Math. Statist., 21, 461, Abs. No. 2.

JOHNSON, N. L. (1958). The mean deviation, with special reference to samples from a Pearson Type III population. Biometrika, 45, 478-483.

KENDALL, M. G. (1956). Exercises in Theoretical Statistics, 2nd. ed. Charles Griffin and Co., Ltd., London.

KENDALL, M. G. and STUART, A. (1958). The Advanced Theory of Statistics, I. Charles Griffin and Co., Ltd., London.

KONDO, T. (1930). A theory of the sampling distribution of standard deviations. Biometrika, 22, 36-64.

LAUE, R. V. (1965). Heterogeneity of variances in a one-way and two-way classification model. Unpublished Ph.D. Thesis, Rutgers - The State University.

LEHMANN, E. L. (1966). Some concepts of dependence. Ann. Math. Statist., 37, 1137-1153.

LE ROUX, J. M. (1931). A study of the distribution of the variance in small samples. Biometrika, 23, 134-190.

LESLIE, R. T. and BROWN, B. M. (1966). Use of range in testing heterogeneity of variance. Biometrika, 53, 221-227.

LIEBLEIN, J. (1953). On the exact evaluation of the variances and covariances of order statistics in samples from the extreme-value distribution. Ann. Math. Statist., 24, 282-287.

MCKAY, A. T. and PEARSON, E. S. (1933). A note on the distribution of range in samples of n. Biometrika, 25, 415-420.

MOORE, P. G. (1957). Normality in quality control charts. Applied Statistics, 6, 171-179.

MORIGUTI, S. (1954). Bounds for second moments of the sample range. Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs., 3, 57-64.

MOSES, L. E. (1956). Some theoretical aspects of the lot plot sampling inspection plan. J. Amer. Statist. Asso., 51, 84-107.

NEYMAN, J. (1941). On a statistical problem arising in routine analyses and in sampling inspections of mass production. Ann. Math. Statist., 12, 46-76.

NEYMAN, J. and PEARSON, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference, I. Biometrika, 20A, 175-240.

NEYMAN, J. and PEARSON, E. S. (1936). Contributions to the theory of testing statistical hypotheses. Statist. Res. Memoirs, I, 1-37.

NORRIS, N. (1935). Inequalities among averages. Ann. Math. Statist., 6, 27-29.

ODEH, R. E. and OLDS, E. G. (1959). Notes on the analysis of variance of logarithms of variances. WADC Tech. Note 59-82, Wright-Patterson AFB, Ohio.

OSTLE, B.; PRAIRIE, R. R.; WIESEN, J. H. (1961). Distribution of the range and midrange when sampling from a negatively skewed right triangle population. Ind. Quality Control, 17, No. 9, 13-15.

PEARSON, E. S. (1929). Note on Dr. Craig's paper. Biometrika, 21, 294-302.

PEARSON, E. S. (1950). Some notes on the use of the range. Biometrika, 37, 88-92.

PEARSON, E. S. (1963). Some problems arising in approximating to probability distributions, using moments. Biometrika, 50, 95-111.

PEARSON, E. S. (1966). Alternative tests for heterogeneity of variance; some Monte Carlo results. Biometrika, 50, 229-234.

PEARSON, E. S. and ADYANTHAYA, N. K. (1928). The distribution of frequency constants in small samples from symmetrical populations. Biometrika, 20A, 356-360.

PEARSON, E. S. and HAINES, J. (1935) The use of the range in place of standard deviation in small samples. J. Roy. Statist. Soc., Suppl. 2, 83-98.

PEARSON, E. S. and HARTLEY, H. O., ed. (1954). Biometrika Tables for Statisticians, Vol. I. Cambridge Univ. Press, Cambridge, England.

PEARSON, E. S. and TUKEY, J. W. (1965). Approximate means and standard deviations based on distances between percentage points of frequency curves. Biometrika, 52, Parts 3 and 4, 533-546.

PITMAN, E. J. G. (1948). Non-Parametric Statistical Inference. Univ. of North Carolina Inst. of Statist. (mimeographed lecture notes).

PLACKETT, R. L. (1947). Limits of the ratio of mean range to standard deviation. Biometrika, 34, 120-122.

RIDER, P. R. (1929). On the distribution of the ratio of mean to standard deviation in small samples from non-normal universes. Biometrika, 21, 124-143.

RIDER, P. R. (1951). The distribution of the range in samples from a discrete rectangular population. J. Amer. Statist. Asso., 46, 375-378.

RIETZ, H. L. (1931). Note on the distribution of the standard deviation of sets of three variates drawn at random from a rectangular distribution. Biometrika, 23, 424-426.

SARHAN, A. E. (1954). Estimation of the mean and standard deviation by order statistics. Ann. Math. Statist., 25, 317-328.

SARHAN, A. E. and GREENBERG, B. G., ed. (1962). Contributions to Order Statistics. John Wiley and Sons, New York.

SCHEFFE, H. (1959). The Analysis of Variance. John Wiley and Sons, New York.

SHAH, B. K. (1966). On the bivariate moments of order statistics from a logistic distribution. Ann. Math. Statist., 37, 1002-1010.

SHAPIRO, S. S. and WILK, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52, Parts 3 and 4, 591-611.

SHONE, K. J. (1949). Relations between the standard deviation and the distribution of range in non-normal populations. J. Roy. Statist. Soc., B, 11, 85-88.

SIOTANI, M. (1955). The significance of the discordant variance estimates. Ann. Inst. Statist. Math., 7, 39-55.

SPRINGER, M. D. (1950). Joint sampling distribution of mean and standard deviation for a chi-square universe. Abstract of Thesis, Univ. of Illinois, 16 pages.

SPRINGER, M. D. (1953). Joint sampling distribution of the mean and standard deviation for probability density function of doubly infinite range. Ann. Math. Statist., 24, 118-122.

STECK, G. P. (1958). A table for computing trivariate normal probabilities. Ann. Math. Statist., 29, 780-800.

STEVENS, W. L. (1936). Hetrogeneity of a set of variances. J. Genetics, 33, 398-399.

TAGUTI, G. (1951). On bias of sample mean and sample variance due to rounding or grouping. Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs., 1, No. 2, 9-14.

TARTER, M. E. and CLARK, V. A. (1965). Properties of the median and other order statistics of logistic variates. Ann. Math. Statist., 36, 1779-1786.

TIPPETT, L. H. C. (1925). On the extreme individuals and the range of samples taken from a normal population. Biometrika, 17, 364-387.

TRUAX, D. R. (1953). An optimum slippage test for the variance of k normal distributions. Ann. Math. Statist., 24, 669-673.

TRUKSA, L. (1940). The simultaneous distribution of mean and standard deviation, and of mean and variance. Biometrika, 31, 256-271.

TUKEY, J. W. (1948). Some elementary problems of importance to small sample practice. Human Biology, 20, 205-214.

TUKEY, J. W. (1960). A survey of sampling from contaminated distributions. Contributions to Probability and Statistics, A volume dedicated to Harold Hotelling, Stanford Univ. Press.

TUKEY, J. W. (1960a). The practical relationship between the common transformations of percentages or fractions and of amounts. Princeton, Statist. Res. Group, Tech. Rept. 36.

TUKEY, J. W. (1962). The future of data analysis. Ann. Math. Statist., 33, 1-67.

VAN DYKE, J. (1961). Numerical investigation of the random variable $y = C(u^{\lambda} - (1-u)^{\lambda})$. Unpublished Working Paper, National Bureau of Standards.

WALD, A. (1939). Contributions to the theory of statistical estimation and testing hypotheses. Ann. Math. Statist., 10, 299-326.

WALSH, J. E. (1956). Validity of approximate normality values for $\mu \pm k\sigma$ areas of practical type continuous populations. Ann. Inst. of Statist. Math., 8, No. 2, 79-86.

APPENDIX A: APPROXIMATIONS TO THE DISTRIBUTION OF $\Sigma\chi_\nu$

Introduction

Let χ_ν denote a random variable distributed as chi, the positive square root of a chi-square variate. That is, let the distribution of χ_ν be given by

$$f_\nu(x) = \frac{x^{\nu-1}e^{-\frac{1}{2}x^2}}{\Gamma(\frac{1}{2}\nu)2^{\frac{1}{2}\nu-1}} \quad .$$

In this appendix we consider approximations to the distribution of

$$T = \sum_{j=1}^{m} \chi_{\nu,j}/\sqrt{\nu} \quad ,$$

where the $\chi_{\nu,j}$ are independently distributed as chi, each with ν degrees of freedom.

Four approximations are compared and it is shown that the approximation based on a scaled chi-square raised to a fractional power provides very good accuracy. It is also shown that the simple approximation provided by a scaled chi with the same expectation and variance, provides fair accuracy for values of m and ν that are not too small. These approximations are used in the computation of approximate percentage points for Eisenhart's test for homogeneity of variance discussed in Section 6.1.

All four approximations are based on Pearson curves with the same lower moments as T, since experience has shown (see for example, Pearson 1963) that this general method often provides useful approximations to the distributions encountered in practice. The basis for the particular distributions chosen as approximations, is outlined below.

Moments

Let $\mu_\nu = E(\chi_\nu / \sqrt{\nu}) = \sqrt{2/\nu}\,\Gamma\left(\frac{\nu+1}{2}\right) / \Gamma\left(\frac{\nu}{2}\right)$.

Then the lower moments of $T = \sum_{i=1}^{m} (\chi_\nu / \sqrt{\nu}$

are given by:

$$E(T) = m\mu_\nu$$

$$E(T^2) = m\left[1 + (m-1)\mu_\nu^2\right]$$

$$V(T) = m(1 - \mu_\nu^2)$$

$$E(T^3) = m\mu_\nu\left[3m - 2 + 1/\nu + (m-1)(m-2)\mu_\nu^2\right]$$

$$\sqrt{\beta_1(T)} = \frac{\mu_\nu}{\sqrt{m}}\left[\frac{1/\nu - 2 + 2\mu_\nu^2}{(1 - \mu_\nu^2)^{3/2}}\right]$$

$$V(T^2) = 2m\left[1/\nu + (m-1)(1 + 2(m - 2 + 1/\nu)\mu_\nu^2 - (2m - 3)\mu_\nu^4)\right]$$

Equate First Two Moments of T and $H\chi_{\xi} / \sqrt{\xi}$

In the first approximation, the distribution of T is approximated by that of a scaled chi with the same first two moments. The coefficients for this approximation are determined by equating the first two moments of T and those of $H\chi_{\xi} / \sqrt{\xi}$. This method requires the solution for H and ξ of the pair of simultaneous equations

$$m\mu_{\nu} = H\mu_{\xi}$$

and

$$H^2 = m + m(m-1)\mu_{\nu}^2 \; .$$

The second equation may be solved for H explicitly then the implicit equation

$$\mu_{\xi} = m\mu_{\nu} / H$$

must be solved for ξ. Here ξ was found by Lagrangian interpolation.

Equate First Two Moments of T^2 and $H^2\chi^2_\eta$

The second approximation is also based on the use of a scaled chi, this time with the same first two moments as T^2. It is the easiest approximation to use since solutions may be obtained in closed form. Equating $E(T^2)$ and $V(T^2)$ to $E(H^2\chi^2_\eta)$ and $V(H^2\chi^2_\eta)$ and solving for H and η yields

$$H^2 = m + m(m-1)\mu^2_\nu$$

and

$$\eta = 2H^4 \;/\; V(T^2).$$

The value of H is the same for both approximations.

Equate First Three Moments of T and $K(\chi_\gamma)^p$

Cadwell (1953) has found that the distributions of a number of measures of dispersion can be approximated with improved accuracy by the distribution of a fractional power of a scaled chi-square with the same first three moments. Let the moments of $(\chi_\gamma)^p$ be denoted by

$$\mu'_\ell = E\left[(\chi_\gamma)^p\right] = \frac{2^{\frac{1}{2}p\ell}\,\Gamma\left(\frac{p\ell + \gamma}{2}\right)}{\Gamma\left(\frac{\gamma}{2}\right)} \quad .$$

Then the use of this method requires the solution of a pair of non-linear equations in two unknowns, p and γ. The equations suggested by Cadwell are

$$CV^2(T) + 1 = \frac{\Gamma(2p + \frac{1}{2}\gamma)\Gamma(\frac{1}{2}\gamma)}{\Gamma^2(p + \frac{1}{2}\gamma)}$$

and

$$\sqrt{\beta_1(T)} \cdot CV^3(T) + 3CV^2(T) + 1 = \frac{\Gamma(3p + \frac{1}{2}\gamma)\Gamma^2(\frac{1}{2}\gamma)}{\Gamma^3(p + \frac{1}{2}\gamma)}$$

where

$$CV^2(T) = V(T) / E^2(T) .$$

A FORTRAN program was written to solve this pair of equations and the coefficients obtained are outlined in Table A.1. The constant K was then obtained by identifying expectations of T and $K(\chi_\gamma)^p$.

Equate First Three Moments of T and Type I Fixed-Start Curve

The (β_1,β_2) points of T fall in the Pearson Type I region (see Figure A.1) so this type of curve was also considered. The Type I fixed-start three-moment curve was used rather than the Type I four-moment curve with floating terminals, since experience has shown that the fixed-start curve generally provides a better fit when the probability in the left tail is not neglegibly small. (See editorial footnote, p.143 to Le Roux, 1931.)

Let the Type I curve be defined by

$$g(y) = Cy^{\alpha-1}(b-y)^{\beta-1} \qquad 0 \le y \le b .$$

The following solution for the parameters α, β, and b was given by Le Roux (1931, p.143).

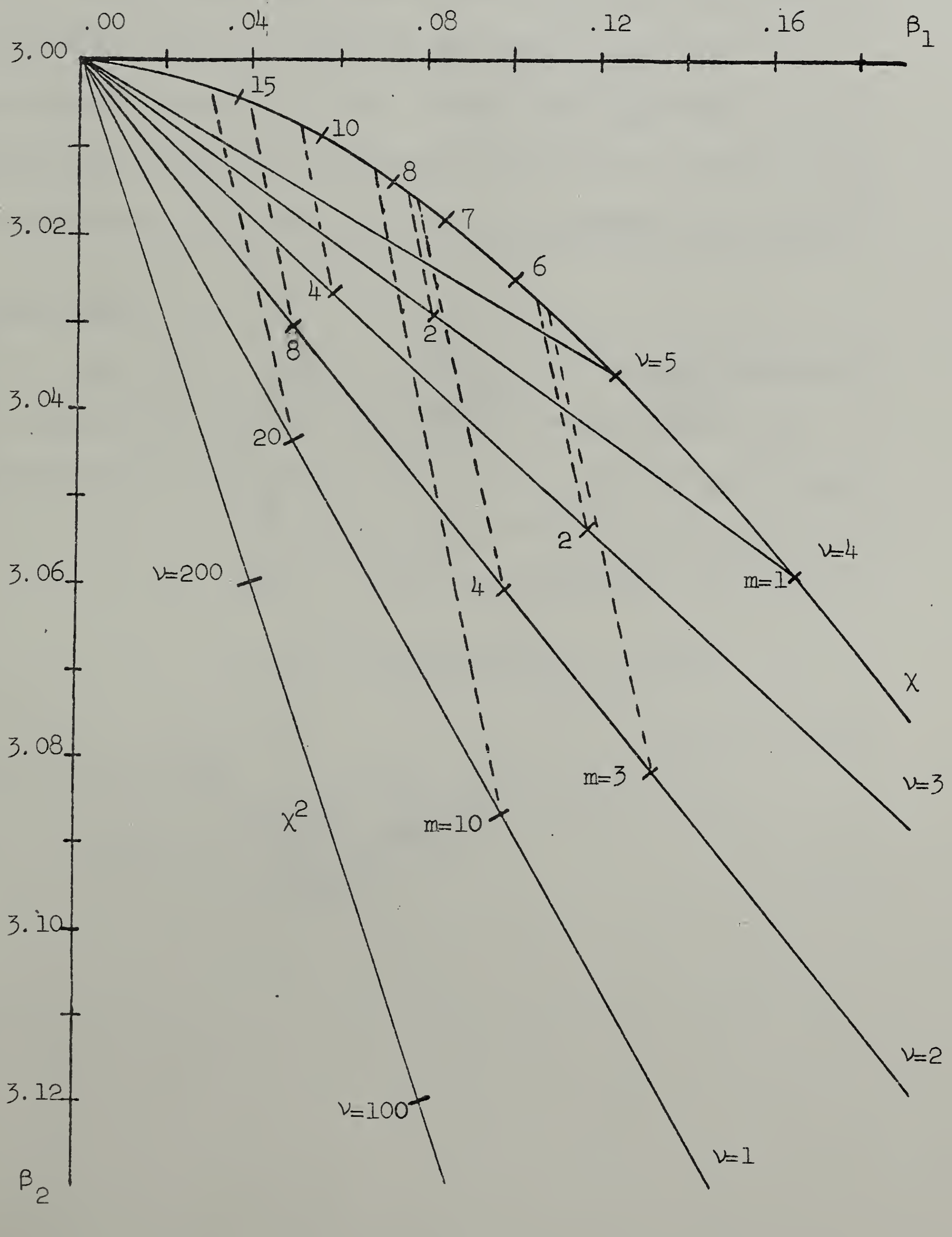
Table A.1

Coefficients for the Cadwell approximation to T. The upper entry is p and the lower is γ.

m	ν = 1	2	3	4	5	6	7	8	10	20
2	.54526 2.22759	.52949 4.28754	.52102 6.29897	.51610 8.2987	.51298 10.2956	.51085 12.2928		.50810 16.2852	.50646 20.2801	.50321 40.275
3	.56161 3.46663	.53921 6.57205		.52128 12.588		.51432 18.5754		.51073 24.5632	.50856 30.5538	
4	.55989 4.70663	.54399 8.85399			.51920 20.8664		.51375 28.8464		.50960 40.825	
5	.57486 5.94624									
6	.57816 7.18565	.54870 13.4142	.53441 19.4515	.52633 25.4464		.51774 37.4161				
8	.58227 9.65333			.52758 34.0153						
10	.58471 12.1401	.55240 22.5292			.52283 52.5587				.51145 102.450	
20	.58955 24.5196	.55514 45.3096			.52402 105.371				.51206 205.154	

Figure A.1

Beta diagram showing the basis for the approximations. The points for χ fall on the curved line while those for χ^2 fall on the straight line on the lower left. The points for $\Sigma\chi_\nu$ fall along the line connecting the point for a single χ_ν with the normal point (0,3). The dashed lines connect the points of the scaled chi used in the first approximation and those of the $\Sigma\chi$ being approximated. (The second chi approximation falls within plotting error of the first.) The Type I area lies to the right of the χ^2 line.



Let $\Psi_1 = E(T^2) / E^2(T)$

and $\Psi_2 = E(T^3) / [E(T) \cdot E(T^2)]$.

Then the solution is given by

$$\alpha = \frac{2(\Psi_1 - \Psi_2)}{2\Psi_2 - \Psi_1\Psi_2 - \Psi_1} ,$$

$$\beta = \frac{2(\Psi_2 - \Psi_1)}{2\Psi_1 - \Psi_2 - 1} - \alpha ,$$

and

$$b = \frac{\alpha + \beta}{\alpha} E(T) .$$

Some Exact Results

For certain values of m and ν it is possible to obtain rather simple forms for $P(T \leq t)$. These results are outlined below.

$\nu = 1, m = 2$ Bland, et al. (1966)

$$P(T \leq t) = \left[2\Phi(t/\sqrt{2}) - 1\right]^2$$

where as always

$$\Phi(y) = \int_{-\infty}^{y} (2\pi)^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}x^2\right\} dx .$$

<u>$\nu = 1,\ m = 3$</u> Bland, et al. (1966)

$$P(T \leq t) = 8\Phi(t/\sqrt{3}\,) - 1 - 48S(t/\sqrt{3},\ 1/\sqrt{2},\ \sqrt{3}\,)$$

where $S(h,a,b)$ is a function related to the trivariate normal distribution. This function has been tabulated by Steck (1958) but the results given here based on the S function were actually independently computed using the same method that Steck used in the original computation of his tables. This procedure obviated the need for three-way interpretation which would otherwise have been necessary. His tables were used as a check on the program.

<u>$m = 2,\ \nu$ even</u>

$$P(T \leq t) = P_{2n}(2a^2) - \frac{e^{-\frac{1}{2}a^2}}{\Gamma(n)2^{2n-1}} \sum_{j=0}^{n-1} \frac{1}{j!\,2^{2j}}$$

$$\cdot \sum_{k=0}^{2j} \binom{2j}{k} (-1)^k a^{4j-2k} \sum_{\ell=0}^{n-j-1} \binom{2n-2j-1}{2\ell} a^{2n-2j-2\ell-1}$$

$$\cdot\, \Gamma\left(k + \ell + \tfrac{1}{2}\right) 2^{k+\ell+\frac{1}{2}}\, P_{2k+2\ell+1}(a^2)$$

where $\quad a = t\sqrt{\nu/2}$

and $\quad P_m(x^2) = \int_0^{x^2} \frac{y^{\frac{1}{2}m-1}\, e^{-\frac{1}{2}y}}{\Gamma(\frac{1}{2}m)2^{\frac{1}{2}m}}\, dy\ .$

m one larger than that for which closed form is available

When the c.d.f. is known for m-1, say $H(z)$ where

$$Z = \sum_{j=1}^{m-1} \chi_\nu / \sqrt{\nu} \quad ,$$

the convolution formula

$$P(T \leq t) = \int_0^t H(t-y)g(y)dy$$

may be integrated by quadrature to obtain results for m. Here $g(y)$ denotes the p.d.f. of $\chi_\nu / \sqrt{\nu}$. Theoretically this procedure could be carried on indefinitely, but practical limits are soon reached.

Computations

The first four moments of T and of the approximations were computed as well as several values of the cumulative distribution function. Let $\mu = E(T)$ and $\sigma^2 = V(T)$, then $P(T \leq \mu + a\sigma)$ was computed for a = 1.5, -1.0, -.5, 0, 1, 2 for each of the four approximations and for the true distribution where possible. The results of these computations are summarized in the tables below.

Error Analysis for Cadwell Approximation

Computation of exact values of the cumulative distribution function was feasible only in a few cases so in the evaluation of the overall goodness of the Cadwell approximation we have had to resort to a comparison of β_2 differences and an appeal to the central limit effect. The following paragraphs outline the basis for our conclusion that the Cadwell approximation improves with m and gives at least four decimal place accuracy for most of the values of m and ν considered.

Pearson (1963) has observed that when the first four moments of an approximation are set equal to the corresponding true moments, "the values of the standardized 5'th and 6'th moments do provide some guide to the degree of correspondence" of the two. Since the moments become less reliable indicators as the order of moment increases, it may be argued that if only the first two (or three) moments of an approximation are identified, the values of the standardized 3'rd and 4'th moments should provide an even better guide to the degree of correspondence. In the Cadwell approximation the first three moments were identified and the standardized fourth moment, β_2, was used as a guide. The computed values of β_2 fell within .04 percent of the true values for $\nu = 2$ through 10 and m less than 20. Most of this difference is due to roundoff error and one would be forced to resort to double precision methods to reduce this figure significantly. For $\nu = 1$

the β_2 differences were somewhat larger but decreased monotonically with m, as is shown in Table A.2. There is reason to believe that due to the central limit effect, the quality of the fit should improve according to $\sqrt{m}$, as it appears to do when measured by the β_2 differences in Table A.2.

Table A.2

Difference between the value of β_2 for the Cadwell approximation and the true value, for $\nu = 1$. The value for the approximation was always the smaller.

m	2	3	4	5	10	20
Absolute	.014	.011	.009	.007	.003	.001
Percent	.40	.34	.27	.22	.11	.05

It is more difficult to directly evaluate the error in the approximation to the cumulative distribution function. However Table A.3a does show that where the exact distribution of T is known, the fit does improve with increasing m and ν and at about the same rate as indicated by the β_2 differences for $\nu = 1$.

On the basis of the above information it seems reasonable to proceed on the assumption that the Cadwell approximation improves, or at least does not get worse, with increasing m. If this assumption is justified, the Cadwell approximation is sufficiently accurate to serve as a basis for an evaluation of the three other less accurate approximations included in this appendix.

Table A.3

Error (X1000) in approximation to the $P(T \leq \mu + a\sigma)$

(a) Cadwell approximation

	approx. ν=	1			2		4		6		8	10
a	prob.	m= 2	3	4	2	3	2	3	2	3	2	2
-1.5	.05	-.6	-.3	-.1	-.0	.0	.0	.0	.0	.0	.0	.0
-1.0	.15	.1	.3	.3	.1	.1	-.0	-.0	-.0	-.0	-.0	-.0
-0.5	.35	.7	.5	.4	.1	.0	-.0	-.0	-.0	-.0	-.0	-.0
0.0	.55	.4	.1	.0	-.1	-.1	-.0	-.0	-.0	-.0	-.0	-.0
1.0	.85	-.4	-.3	-.2	-.0	-.0	.0	.0	.0	.0	.0	.0
2.0	.95	-.0	.0	.0	.0	.0	.0	.0	.0	.0	.0	.0

(b) Pearson Type I approximation

-1.5	.05	-4.8	-2.8	-1.6	-1.6	-.7	-.3	-.1	-.1	.0	-.0	.0
-1.0	.15	-.2	1.2	1.4	1.6	1.4	1.2	.9	1.0	.7	.8	.6
-0.5	.35	5.8	4.4	3.4	3.6	2.5	2.0	1.3	1.4	.9	1.0	.8
0.0	.55	4.6	2.8	1.9	2.0	1.1	.8	.4	.4	.2	.3	.2
1.0	.85	-3.0	-2.6	-2.1	-2.4	-1.7	-1.5	-1.0	-1.1	-.7	-.8	-.7
2.0	.95	-1.5	-.8	-.5	-.5	-.2	-.1	-.0	.0	.0	.0	.0

Table A.3 (Continued)

(c) Chi(T) approximation

a	approx. prob.	ν = 1, m = 2	ν = 1, m = 3	ν = 1, m = 4	ν = 2, m = 2	ν = 2, m = 3	ν = 4, m = 2	ν = 4, m = 3	ν = 6, m = 2	ν = 6, m = 3	ν = 8, m = 2	ν = 10, m = 2
-1.5	.05	2.6	3.6	3.6	1.6	1.7	.6	.6	.3	.3	.2	.1
-1.0	.15	3.2	3.0	2.5	1.0	.8	.2	.2	.1	.1	.0	.0
-0.5	.35	-0.2	-1.3	-1.9	-0.9	-1.3	-0.6	-0.7	-0.4	-0.4	-0.2	-0.2
.0	.55	-3.4	-4.7	-5.0	-2.3	-2.6	-1.0	-1.1	-0.6	-0.6	-0.4	-0.3
1.0	.85	-2.0	-1.9	-1.6	-0.6	-0.5	-0.1	-0.1	-0.0	-0.0	-0.0	-0.0
2.0	.95	1.0	1.5	1.6	.8	.9	0.4	.4	.2	.2	.1	.1

(d) Chi(T^2) approximation

a	approx. prob.	ν = 1, m = 2	ν = 1, m = 3	ν = 1, m = 4	ν = 2, m = 2	ν = 2, m = 3	ν = 4, m = 2	ν = 4, m = 3	ν = 6, m = 2	ν = 6, m = 3	ν = 8, m = 2	ν = 10, m = 2
-1.5	.05	5.3	6.8	6.7	2.9	2.9	1.0	1.0	.5	.5	.3	.2
-1.0	.15	9.0	8.6	7.3	3.1	2.7	.8	.7	.3	.3	.2	.1
-0.5	.35	5.3	3.9	2.6	1.0	.4	-.1	-.2	-.1	-.2	-.1	-.1
0.0	.55	-.3	-1.9	-2.8	-1.4	-1.9	-.8	-.9	-.5	-.6	-.3	-.2
1.0	.85	-3.1	-3.4	-3.2	-1.3	-1.3	-.4	-.4	-.2	-.2	-.1	-.1
2.0	.95	-.3	.1	.3	.2	.4	.2	.2	.1	.2	.1	.1

Error Analysis for Type I Approximation

It was originally hoped that the Pearson Type I curve would provide a sufficiently accurate approximation to serve as a basis for the evaluation of the chi approximations. However, such was not the case. This approximation actually gave errors of roughly the same order of magnitude as the chi approximations. Tables A.3b and A.4 give an indication of the accuracy of this three-moment approximation.

Error Analysis for the Scaled Chi Approximations

Two chi approximations were considered, one, denoted by chi(T), in which the first two moments of T are identified with the corresponding moments of the approximation, and the other, denoted by chi(T^2), in which the first two moments of T^2 are identified. The former proved to be slightly more accurate especially in the lower tail which is important in the application considered in Section 6.1. Tables A.3c, A.3d, A.5, and A.6 give an indication of the accuracy of these two approximations.

Table A.4

Difference (X1000) between the Pearson Type I and Cadwell approximations for the $P(T \leq \mu + a\sigma)$.

m	ν	a = -1.5	-1	-.5	0	1	2
5	1	-1.0	1.1	2.5	1.4	-1.7	-0.3
10	1	-0.2	.8	1.3	.5	-0.9	-0.1
20	1	-0.0	.5	.6	.2	-0.5	.0
4	2	-0.4	1.1	1.8	.8	-1.4	-0.1
10	2	-0.0	.5	.7	.2	-0.6	.0
20	2	.0	.3	.3	.1	-0.3	.0
2	3	-0.7	1.4	2.6	1.2	-1.9	-0.2
8	4	.0	.4	.5	.1	-0.4	.0
2	5	-0.2	1.1	1.6	.6	-1.3	-0.0
4	5	.0	.6	.8	.2	-0.7	.0
10	5	.0	.3	.3	.1	-0.3	.0
20	5	.0	.1	.2	.0	-0.1	.0
4	10	.0	.3	.4	.1	-0.4	.0
10	10	.0	.1	.2	.0	-0.1	.0

Table A.5

Difference (X1000) between the chi(T) and Cadwell approximations for the $P(T \leq \mu + a\sigma)$

		a					
m	ν	-1.5	-1	-.5	0	1	2
5	1	3.5	1.8	-2.4	-5.0	-1.2	1.7
10	1	2.6	.9	-2.4	-4.3	-0.7	1.5
20	1	1.8	.5	-2.0	-3.3	-0.4	1.2
4	2	1.6	.6	-1.4	-2.6	-0.4	.9
10	2	1.1	.3	-1.2	-2.0	-0.2	.8
20	2	.7	.1	-0.9	-1.5	-0.1	.6
2	3	.9	.4	-0.7	-1.4	-0.3	.5
8	4	.4	.1	-0.5	-0.9	-0.1	.3
2	5	.4	.1	-0.4	-0.7	-0.1	.3
4	5	.4	.1	-0.5	-0.8	-0.1	.3
10	5	.3	.0	-0.4	-0.6	-0.0	.2
20	5	.2	.0	-0.3	-0.4	-0.0	.2
4	10	.2	.1	-0.1	-0.3	-0.1	.1
10	10	.1	.0	-0.1	-0.2	-0.0	.1

Table A.6

Difference (X1000) between the chi(T^2) and Cadwell approximations for the $P(T \leq \mu + a\sigma)$

		a					
m	ν	-1.5	-1.0	-.5	0	1	2
5	1	6.2	6.0	1.4	-3.2	-2.8	.5
10	1	4.2	3.3	-0.3	-3.5	-1.9	.8
20	1	2.6	1.7	-0.9	-3.0	-1.1	.8
4	2	2.6	2.2	-0.0	-2.0	-1.2	.4
10	2	1.6	1.0	-0.6	-1.8	-0.7	.5
20	2	1.0	.5	-0.6	-1.4	-0.4	.4
2	3	1.6	1.4	.2	-1.0	-0.7	.2
8	4	.6	.3	-0.3	-0.8	-0.2	.2
2	5	.7	.5	-0.1	-0.6	-0.3	.1
4	5	.6	.4	-0.2	-0.7	-0.2	.2
10	5	.4	.2	-0.3	-0.6	-0.1	.2
20	5	.2	.1	-0.2	-0.4	-0.1	.1
4	10	.2	.1	-0.1	-0.3	-0.1	.1
10	10	.1	.0	-0.1	-0.2	-0.0	.1

Discussion

Four approximations to the distribution of $T = \Sigma\chi_\nu / \sqrt{\nu}$ have been compared and it has been shown that the fractional powered scaled chi approximation in which the first three moments are identified with those of T provides a very good fit. The other three approximations are an order of magnitude less accurate. Of these three the Pearson Type I approximation based on a three-moment fit is slightly more accurate but it would seem to be significantly less tractable. For small values of m and ν the chi(T) approximation is slightly more accurate than the chi(T^2) approximation but its parameters are more difficult to compute. Thus the choice between these two chi approximations each based on an identification of two moments, would have to be made on the basis of the needs of a particular problem.

APPENDIX B. ALTERNATIVES TO THE POWER FUNCTION

Abstract

There has recently been a rekindling of interest in non-asymptotic alternatives to the power function as a means of judging the relative goodness of statistical tests. Hogben, Pinkham, and Wilk (1962), Hogben (1963) and Dempster and Schatzoff (1965) have proposed the use of the expected significance level and Geary (1966) has recently proposed the average critical value and the median critical value. The purpose of this appendix is to introduce two more criteria termed the "median significance level" and the "significance level of the average" and to illustrate by means of some simple examples the relationships among the various measures.

Introduction

The power function (Neyman and Pearson, 1936) of a test of $H_0: \theta = \theta_0$ for fixed sample size n, is usually expressed as a function of the parameter θ for fixed size of test α. Let T be a test statistic for testing H_0 such that H_0 is rejected if T is too large and let t_α be such that $P(T \geq t_\alpha | \theta_0) = \alpha$. Then the power function for the test of size α is equal to $P(T \geq t_\alpha | \theta)$. We assume throughout that the test statistic has a continuous distribution and that the set of alternatives is indexed by a scalar parameter θ .

If the power function is thought of as a family of curves, one for each α, then the power function contains "all" of the information about a test. The most important defect in the power function is, however, a practical one. It is usually very difficult to compute the distribution of the test statistic under the alternative hypotheses and a large 2-way (or multi-way) table is required if information about more than one value of α is needed. When θ is a vector parameter the problem is usually compounded (except when the alternative reduces to a single non-centrality parameter as, for example, in certain χ^2 tests).

Asymptotic approaches to the comparison of tests, which do not require the whole power function, have been suggested by Pitman (1948), Chernoff (1952), and Bahadur (1960). These have proven useful in certain contexts, though by their very nature are not fully satisfactory in others. In the framework of decision theory, other approaches to the comparison of tests have been suggested (Wald, 1939) but are beyond the scope of the present work.

Average Critical Value

Geary (1966), continuing in the "Neyman-Pearson" framework, introduced the average critical value (ACV) concept as a simple practical means of comparing tests whose power functions are not known. The ACV of a test for a given α is the value of θ for which the expectation of the test statistic is equal to the critical value under the null hypothesis. That is, for an upper-tailed test of H_0 of size α, the ACV = θ^*, where θ^* is the solution of the implicit equation

(B.1) $$\mu_{\theta^*} = t_\alpha ,$$

where $$\mu_\theta = E(T|\theta) .$$

Figure B.1 illustrates this definition if one imagines the alternative being chosen so that μ_θ is identical to t_α .

The ACV is a function of α and takes on values θ in the parameter set. It is a single-valued function only for tests that are essentially one-sided. It is not uniquely defined for hypothesis testing problems in which the set of alternatives corresponds to a vector parameter.

Geary also suggested another measure which may be called the median critical value (MCV). The MCV is analogously defined to be the value of θ for the alternative under which the median of the test statistic equals the α critical value. In functional notation, the MCV is defined as the solution θ^* of the implicit equation,

(B.2) $$\xi_{\theta^*} = t_\alpha ,$$

where ξ_θ is such that

$$P(T \geq \xi_\theta | \theta) = \frac{1}{2} .$$

This definition is illustrated in figure B.1 if one imagines the alternative being chosen so that ξ_θ is identical to t_α .

Figure B.1

An illustration of the definitions of Geary's "average critical value" and the "median critical value." Imagine the alternative H_θ being chosen so that μ_θ or ξ_θ coincides with t_α . Then that value of θ is the ACV or MCV, respectively.

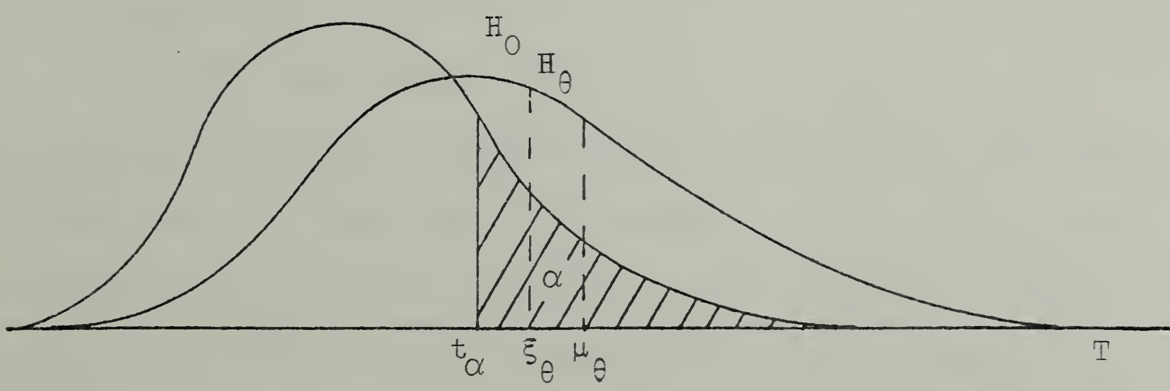
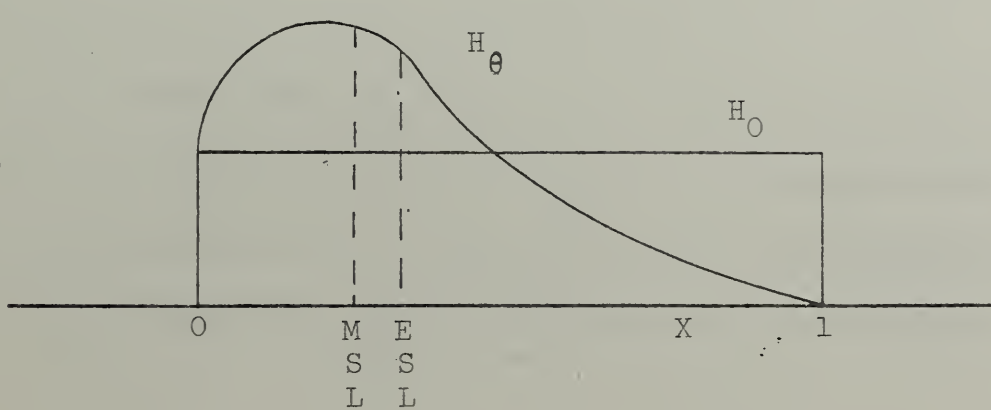


Figure B.2

A hypothetical example showing the distribution of the "observed significance level" under H_0 and under a given H_θ . The values of the "expected significance level" and the "median significance level" are also shown.



Expected Significance Level

For "those statisticians who regard tests of significance as tools for statistical inference rather than procedures for accepting or rejecting hypotheses," Hogben, Pinkham, and Wilk (1962), Hogben (1963), and Dempster and Schatzoff (1965) have suggested that the distribution of the "observed significance level" can serve as a complete alternative to the power function. To quote Hogben; "This point of view, which is independent of a prechosen probability of Type I error, provides a means other than power for describing and assessing the properties of a test of significance."

Let T denote a test statistic with continuous density $f_0(t)$ under the null hypothesis H_0, and suppose that H_0 is such that an unusually large observed value of T would cast doubt upon the validity of H_0. For an observed value of the test statistic T, let $X = \int_T^\infty f_0(t)dt$. Then X is the observed significance level, and is a random variable. Under the null hypothesis, H_0, X is distributed uniformly on (0,1) but under an alternative hypothesis, H_A, smaller values of X are more probable. Figure B.2 illustrates how the distribution of the significance level might look for a given alternative.

Since the distribution of the significance level is difficult to manage and leads to the same problems of tabulation as the power function, Hogben proposed that the expectation and variance of the significance level be used as partial characterizations of its distribution.

Dempster and Schatzoff, following essentially the same argument as Hogben but emphasizing the problem of the choice of a test statistic rather than the assessment of the properties of a single test, note that a comparison of the distributions of the significance levels of two competing tests does not always force the choice of a test statistic. Whereas Hogben had suggested that the first <u>two</u> moments might serve to partially characterize the distribution of the significance level, they suggest that if a single criterion is desired, the first moment or <u>expected significance level</u> (ESL) might be a good choice.

Convenient expressions for the ESL are given by

(B.3) $$ESL = 1 - \int F_0(x) f_\theta(x)dx$$

and

(B.4) $$ESL = P(T_0 \geq T_\theta)$$

where T_0 is distributed according to f_0 and T_θ is independently distributed according to f_θ, the density of the test statistic under the alternative θ. $F_0(x)$ is the cumulative distribution function of T_0. Dempster and Schatzoff also showed that 1 - ESL was equivalent to an unweighted average of the power function over all values of α.

The Median Significance Level

The median significance level (MSL) for a given alternative hypothesis, H_θ , is analogously defined to be the median of the distribution of the significance level under the given alternative. That is, if X is defined as in the previous section and if ζ_θ is such that

$$P(X \leq \zeta_\theta | H_\theta) = \frac{1}{2} \quad ,$$

then the

$$MSL = \zeta_\theta \quad .$$

This definition is illustrated in figure B.2.

The concept of the MSL is a natural outgrowth of the concept of the ESL and seems to have at least one important advantage over the ESL procedure, namely ease of computation. It is shown below that for one-sided tests, computation of the MSL requires a knowledge of the alternative distribution only to the extent of knowing the median of the test statistic. The MSL like the ESL is a function of the parameter θ and the values it assumes are probabilities (significance levels).

One important advantage that the ESL and MSL have over the ACV, MCV and power function is the use of the "built-in-metric" intuitively available to those who are accustomed to thinking in terms of observed significance levels. The hope is that this method of thinking will be more closely akin to that actually used by many practicing statisticians.

The MSL also has all the advantages (and disadvantages) relative to the ESL that the median has relative to the mean as a measure of location of a highly skewed distribution. Since the distribution of the significance level is typically highly skewed for even moderately distant alternatives, this is no small point. Some of the examples below illustrate this quite dramatically. The following theorem is relevant.

<u>THEOREM</u>: If

(I) $$f_0(\xi_\theta - y) \geq f_0(\xi_\theta + y)$$

and

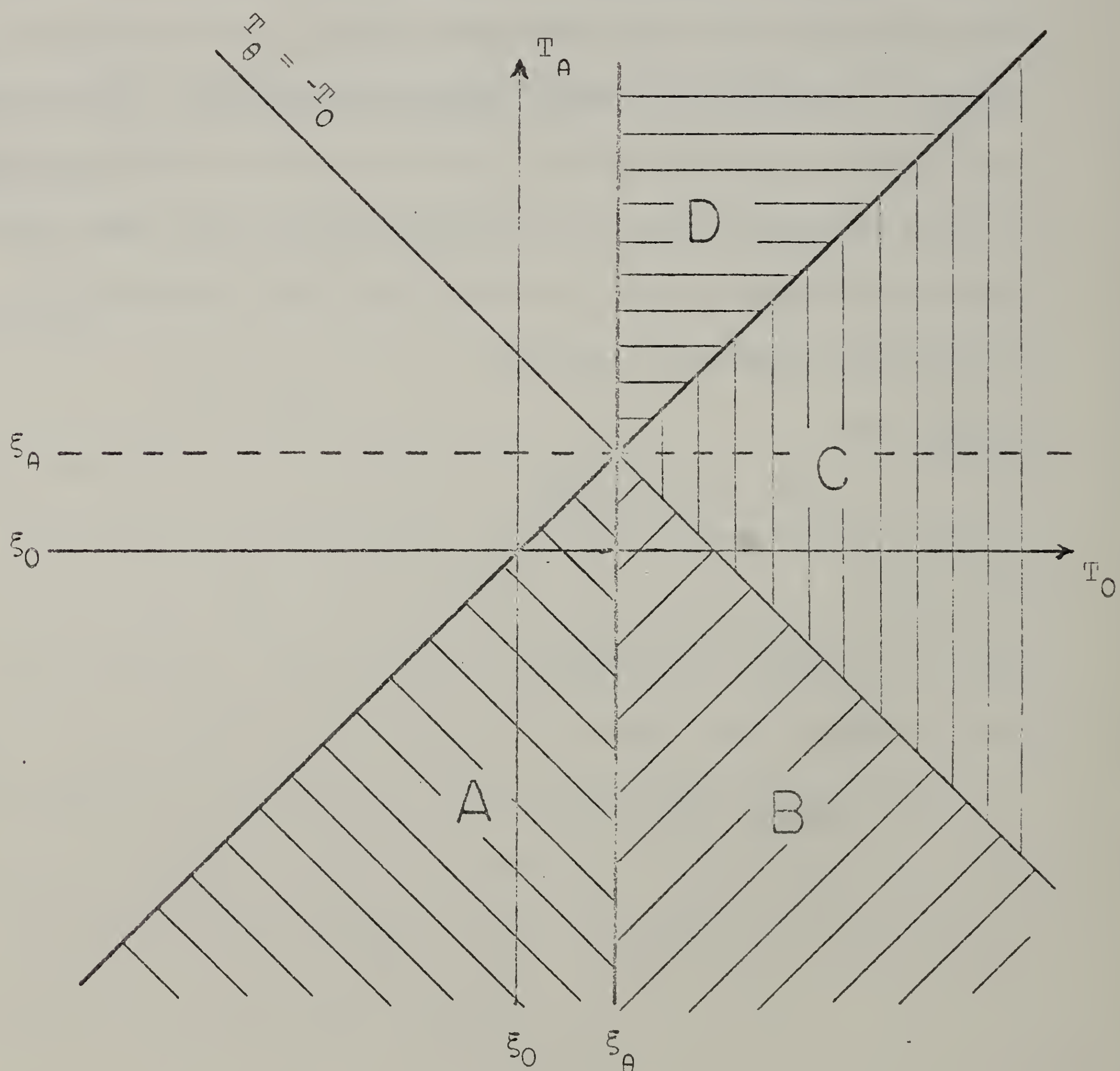
(II) $$f_\theta(\xi_\theta - y) \geq f_\theta(\xi_\theta + y)$$

hold for all $y > 0$, then

$$\mathrm{ESL}(\theta) \geq \mathrm{MSL}(\theta) \quad .$$

Figure B.3

The ESL is greater than the MSL if $P(A) > P(D)$ since the ESL $= P(A) + P(B) + P(C)$ and the MSL $= P(B) + P(C) + P(D)$.



PROOF: The proof is immediate by noting in figure B.3

that $ESL = P(A) + P(B) + P(C)$

and $MSL = P(B) + P(C) + P(D)$,

so $ESL \geq MSL$ if $P(A) \geq P(D)$.

But by condition (II) above, $P(B) \geq P(D)$ and by condition (I), $P(A) \geq P(B)$.

Condition (I) holds, for example, when f_0 is unimodal and symmetric. From the examples it would seem that a much more general result could be established, but this has not yet been done.

The ESL and MSL are completely general in the sense that investigations based on them may be applied to any test, whereas the ACV and MCV are single valued functions only for one-sided tests, and are not uniquely defined for vector-valued parameters. However, one-sided tests of scalar-valued parameters occupy an important position in statistical methods, so this is not a prohibitive constraint.

Defining X, ξ_θ and ζ_θ as before, and continuing to restrict attention to upper-tailed one-sided tests, it follows that

$$MSL = \zeta_\theta = 1 - F_0\,(\xi_\theta) = P(T \geq \xi_\theta | H_0) \ .$$

Thus, the median of the test statistic rather than the median of the significance level may be used in MSL computations, and the more troublesome distribution of the observed significance level can be avoided entirely. This is true for one-sided tests since medians (and other percentiles) are preserved under monotonic transformations, though means are not.

The following computational procedure for the MSL for one-sided tests may be found useful:

1. Specify H_O, H_θ, n, and test statistic T.
2. Find ξ_θ, the median of T under the alternative H_θ.
3. Then the MSL = $P\{T \geq \xi_\theta | H_O\}$ (for lower-tailed test reverse inequality).

It is now seen that the MSL (function), even though arrived at by an entirely different argument, is in reality the inverse function of Geary's MCV (function). That is, the value of the MSL for an alternative θ, is the significance level that has a value of the MCV equal to θ. The implicit equation $\xi_\theta = t_\alpha$ is solved for θ to find the value of the MCV for a given α or solved for α to find the value of the MSL for a given θ.

This observation leads us to suggest the following compromise criterion.

The Significance Level of the Average

The median significance level (function) of the preceding section is seen to be more readily computable than the ESL (function) at least for one-sided tests, while sharing the ESL's principal advantage of the use of the "built-in-metric." However, the MSL (or equivalently the MCV) is still more difficult computationally than the ACV since, as Geary has pointed out, the median of the test statistic under an alternative is usually more difficult to obtain than the mean.

Let the <u>significance level of the average</u> (SLA) be defined as the inverse function of the ACV (function). That is, the implicit equation $\mu_\theta = t_\alpha$ is solved for α to find the value of the SLA for a given θ or solved for θ to find the value of the ACV for a given α. Alternatively, the SLA may be thought of as the significance level attained by a test when its test statistic T assumes the value μ_θ, the expectation of T for a given alternative hypothesis H_θ .

Then the SLA has the computational simplicity of the ACV while making use of the "built-in-metric" of the ESL and MSL. The MSL and SLA are identical for situations where the test statistic has identical mean and median for all θ (e.g., symmetrical distributions) and the MSL is greater than the SLA when the distribution of the test statistic under the alternative is skewed toward the rejection region (e.g., for upper-tailed tests, when $\mu_\theta > \xi_\theta$). Examples are given below showing some relationships among the measures for some common tests.

Remarks

1. The following expressions are interesting to compare:

$$ESL = P(T_0 \geq T_\theta)$$

$$MSL = P(T_0 \geq \xi_\theta)$$

$$SLA = P(T_0 \geq \mu_\theta) \quad .$$

2. The ACV and its inverse function, the SLA, are not invariant under non-linear transformations of the test statistic, though the ESL and MSL are.

3. The MSL may be thought of as answering the question: "at what significance level do I have a 50-50 chance of rejecting the hypothesis if the true situation is as given?" That is, the MSL is the significance level for which there is power ½ at H_θ.

4. Similarly the MCV may be thought of as answering the question: "for which alternative do I have a 50-50 chance of rejecting the hypothesis at the α level?"
That is, the MCV is the alternative (parameter value) for which there is power ½ at level α.

5. The MCV concept is formally the same as Hamaker's (1950) "point of control" or "indifference quality" concept in acceptance sampling. He has proposed that sampling plans be classified according to the percent defective that has probability of acceptance equal to ½.

6. If f_0 is symmetric and f_θ differs only by a location parameter, then 1 - MSL(θ) is equal to the power at θ for $\alpha = .5$.
7. Let $T^{(1)}$ and $T^{(2)}$ denote the test statistics of two competing tests and let $E(T^{(j)}|\theta) \equiv \theta$ for all θ, $j = 1,2$. Then if $t_\alpha^{(1)} \leq t_\alpha^{(2)}$ for all α the ACV (or SLA) criterion will always indicate that the test based on $T^{(1)}$ is superior to the test based on $T^{(2)}$. However, Sundrum (1954) has shown that, under the above conditions, the test based on $T^{(1)}$ is not necessarily always more powerful than the test based on $T^{(2)}$, for all α. If $\xi^{(1)} \equiv \xi^{(2)} \equiv \theta$ for all θ, an analogous result holds for the MCV (or MSL) criterion.
8. Generalization: Any other percentile (say the p'th) of the distribution of the significance level could be used to obtain the "p'th percentile significance level." That is, the significance level that would be exceeded p percent of the time.

Examples

The following examples are given to indicate the general methods involved and as a means of comparing the several criteria. They are stated in terms of the ESL, MSL, and SLA but may be readily interpreted in the ACV and MCV framework as will be illustrated.

Example 1: (Figure B.4)

$f_\tau(x) = \tau^{-1} \exp\{-x/\tau\} \qquad 0 \leq x < \infty$

H_0: $\tau = 1$

H_θ: $\tau = \theta > 1$

Choose n = 1 and upper-tailed test based on X. Then the relations

$$\exp\left\{ -t_\alpha \right\} = \alpha$$

$$\mu_\theta = \theta$$

and

$$\exp\left\{-\xi_\theta/\theta\right\} = \frac{1}{2}$$

define t_α, μ_θ, and ξ_θ . Using (B.1)

and solving $\qquad \theta = -\ln(\alpha)$

for α yields $\qquad SLA = e^{-\theta}$

and for θ yields $\qquad ACV = -\ln(\alpha)$.

Now using (B.2) and solving

$$\theta \ln(2) = -\ln(\alpha)$$

for α yields $\qquad MSL = (\frac{1}{2})^\theta$,

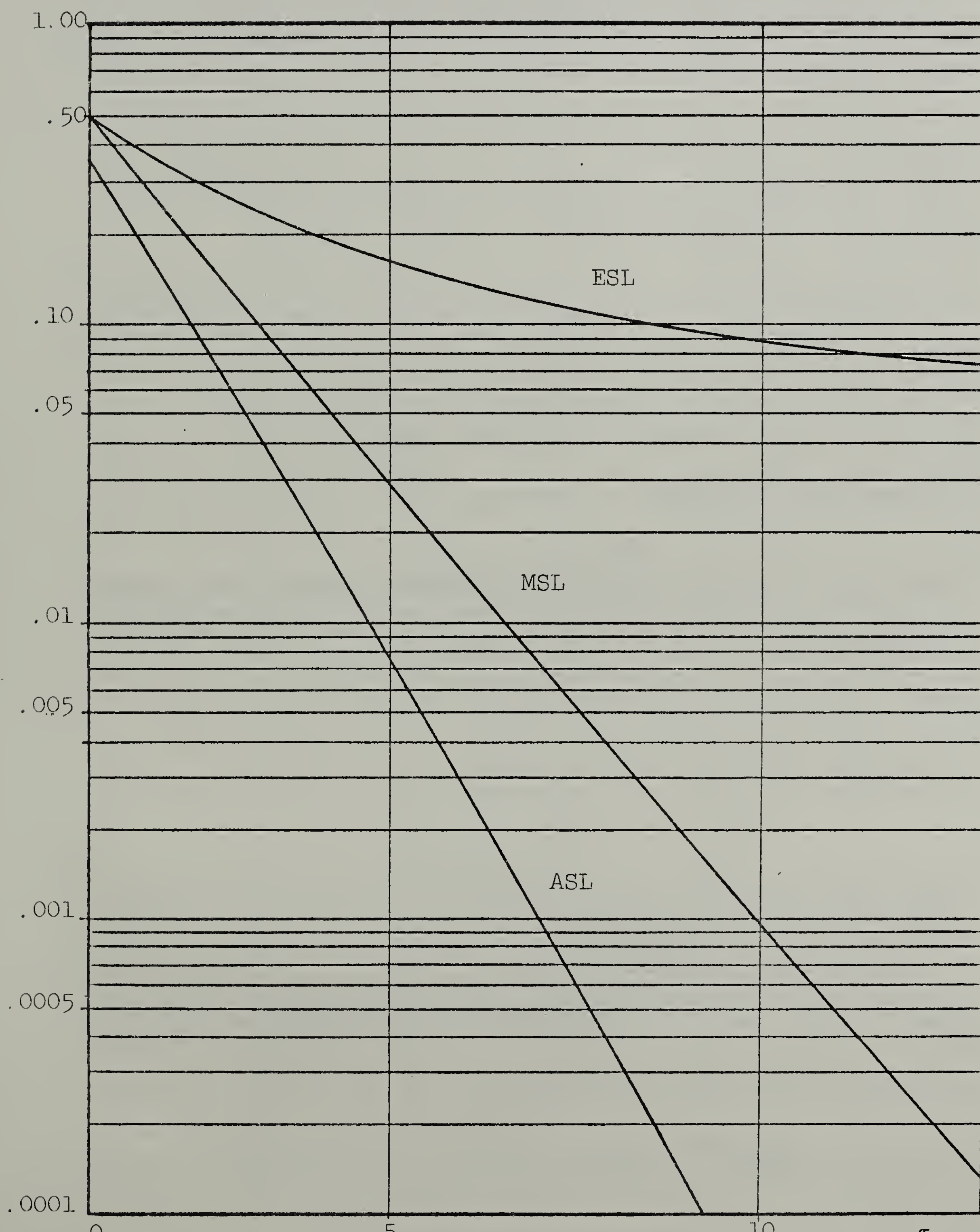
and for θ yields $\qquad MCV = -\ln(\alpha)/\ln(2)$.

Equation (B.4) may be used to find

$$ESL = 1/(1 + \theta) .$$

Figure B.4

Upper tailed test for scale parameter of exponential distribution (n=1).



These results are shown graphically in figure B.4 and may be interpreted in the ESL-MSL-SLA framework by fixing θ and reading the significance level (α) or in the ACV-MCV framework by fixing α and reading θ.

Example 2: (Figure B.5)

$$f_\nu(x) = (2\pi)^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}(x-\nu)^2\right\} \qquad -\infty < x < \infty$$

H_0: $\nu = 0$

H_θ: $\nu = \theta > 0$.

Choose $n = 1$ and an upper-tailed test based on X.

Let $\Phi(x) = \int_{-\infty}^{x} f_0(x)dx$

then $t_\alpha = \Phi^{-1}(1-\alpha)$, where F^{-1} in general denotes the inverse function of F.

Here $\xi_\theta \equiv \mu_\theta = \theta$.

So the MSL $\equiv$ SLA is found by solving

$$\theta = \Phi^{-1}(1-\alpha) \quad \text{for} \quad \alpha.$$

Hence MSL = SLA = $1 - \Phi(\theta) = \Phi(-\theta)$.

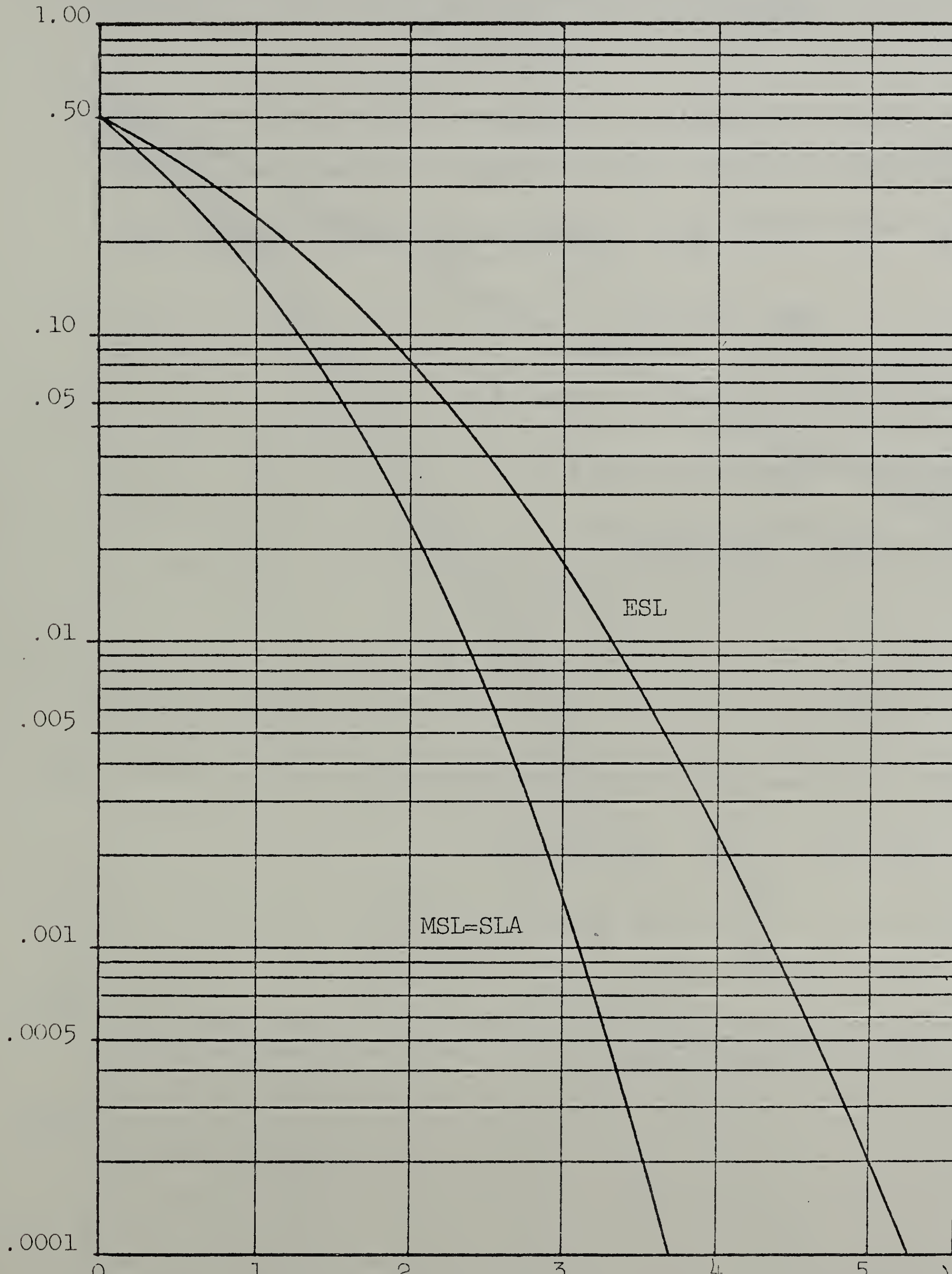
We now need the following well known

Lemma: If X_0 and X_1 are independently distributed as $N(\mu_0, \sigma_0^2)$ and $N(\mu_1, \sigma_1^2)$, i.e., X_i has p.d.f.

$$(2\pi\sigma_i^2)^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\left(\frac{X-\mu_i}{\sigma_i}\right)^2\right\} \qquad -\infty < x < \infty$$

Figure B.5

Upper tailed test for mean of normal distribution (n=1).



then,

$$P(X_0 \geq X_1) = \Phi\left(\frac{\mu_0 - \mu_1}{\sqrt{\sigma_0^2 + \sigma_1^2}}\right).$$

(The lemma is an immediate consequence of the fact that a linear combination of normal random variables is normally distributed.)

Since ESL = $P(X_0 \geq X_\theta)$, we find in the present example that

$$\text{ESL} = \Phi\left(\frac{-\theta}{\sqrt{2}}\right) .$$

These results are shown in figure B.5.

Example 3: (Figures B.6 and B.7)

$$f(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}\left(\frac{x-\nu}{\sigma}\right)^2\right\} \qquad -\infty < x < \infty$$

H_0: $\sigma = 1$

H_θ: $\sigma = \theta > 1$.

Choose n = 5 and upper-tailed tests based on:

(1) $S^2 = \sum_{i=1}^{5} (X_i - \bar{X})^2 / (n-1)$

(2) $S = \sqrt{S^2}$

(3) $R = X_{max} - X_{min}$

Figure B.6
Upper tailed test for σ from normal distribution (n=5)

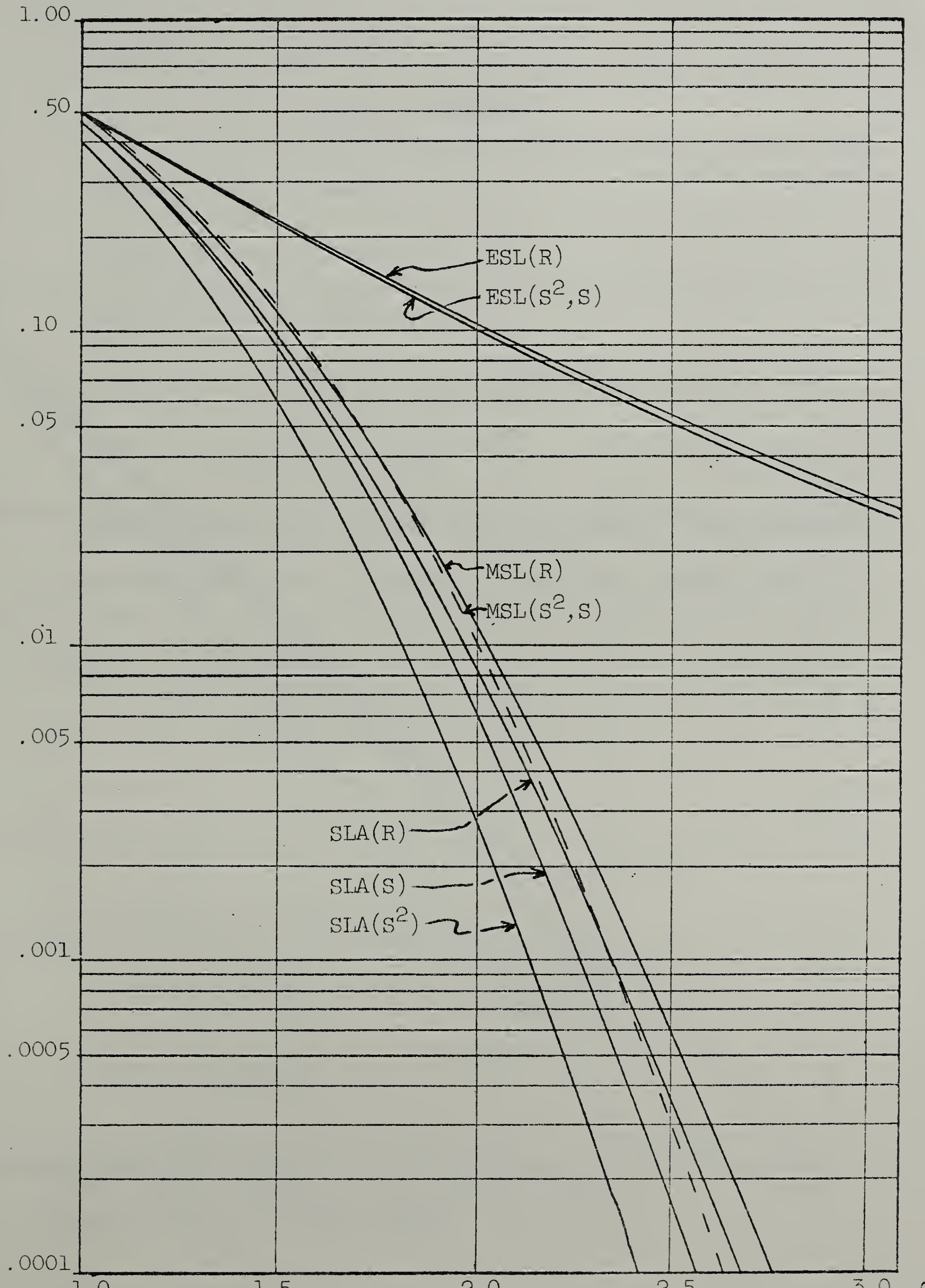
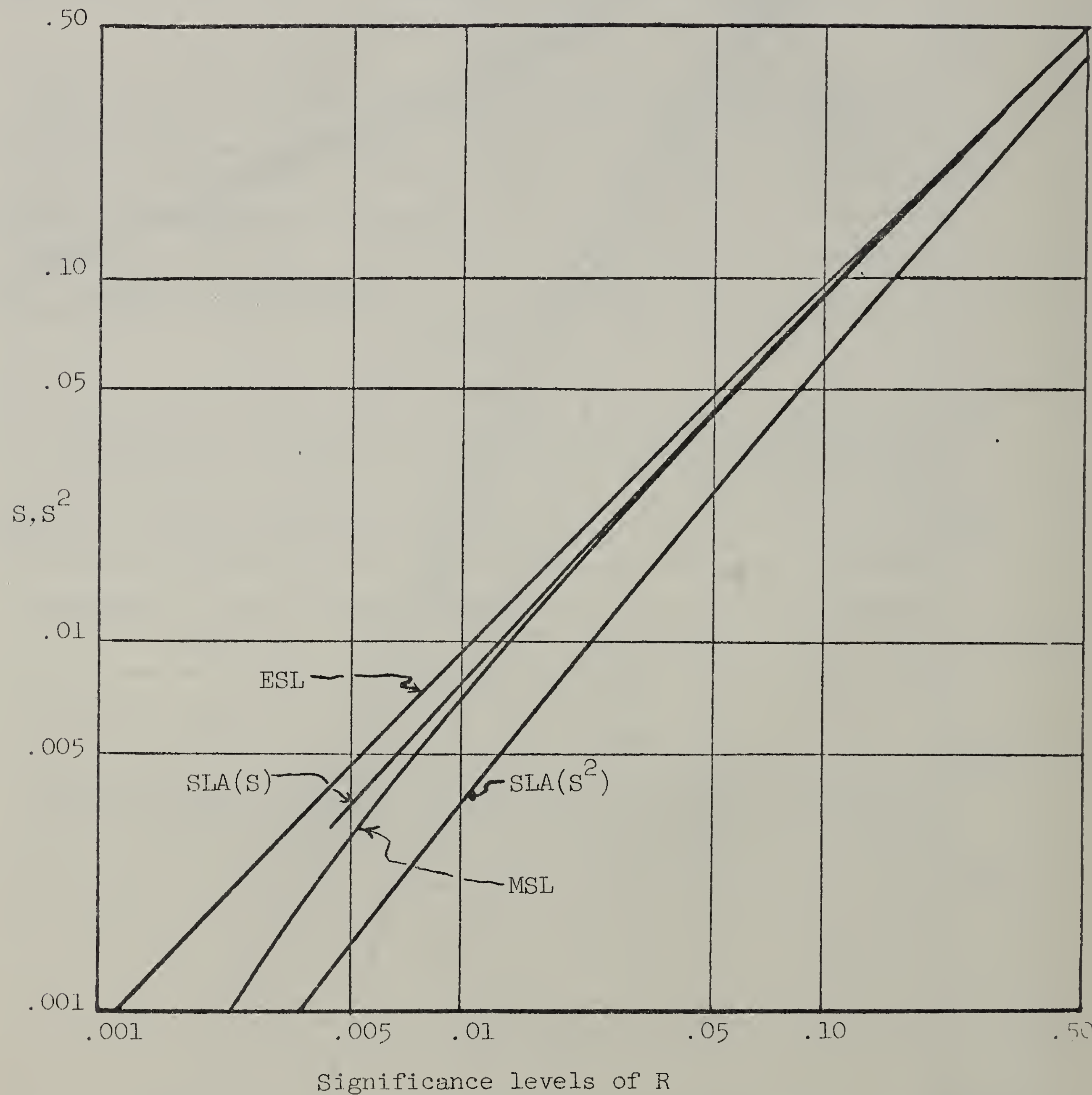


Figure B.7

Upper tailed test for σ from normal distribution (n=5): a comparison of S and R.



Then

$$SLA(S^2) = (1 + 2\theta^2)e^{-2\theta^2}$$

$$SLA(S) = (1 + 1.767\theta^2)e^{-1.767\theta^2}$$

$$SLA(R) = P(R > 2.326\theta | \sigma = 1)^*$$

$$MSL(S^2,S) = (1 + 1.678\theta^2)e^{-1.678\theta^2}$$

$$MSL(R) = P(R > 2.257\theta | \sigma = 1)^*$$

$$ESL(S^2,S) = P(S_0^2 > S_\theta^2) = P(F_{4,4} > \theta^2) = (1 + 3\theta^2)/(1 + \theta^2)^3$$

$$ESL(R) = P(R_0 > R_\theta) = P(\frac{R_1}{R_2} > \theta)^{**} \quad .$$

The different type of plot shown in figure B.7 was obtained by plotting the $ESL(S^2)$ vs. $ESL(R)$; $MSL(S^2)$ vs. $MSL(R)$; etc.

Example 4: (Figure B.8)

$$f_\nu(x) = (2\pi)^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}(x-\nu)^2\right\} \qquad -\infty < x < \infty$$

H_0: $\nu = 0$

H_θ: $\nu = \theta > 0$.

Choose upper-tailed tests based on

(1) $\bar{X} = \sum_{i=1}^{n} X_i/n$

(2) $\tilde{X}$ = median X

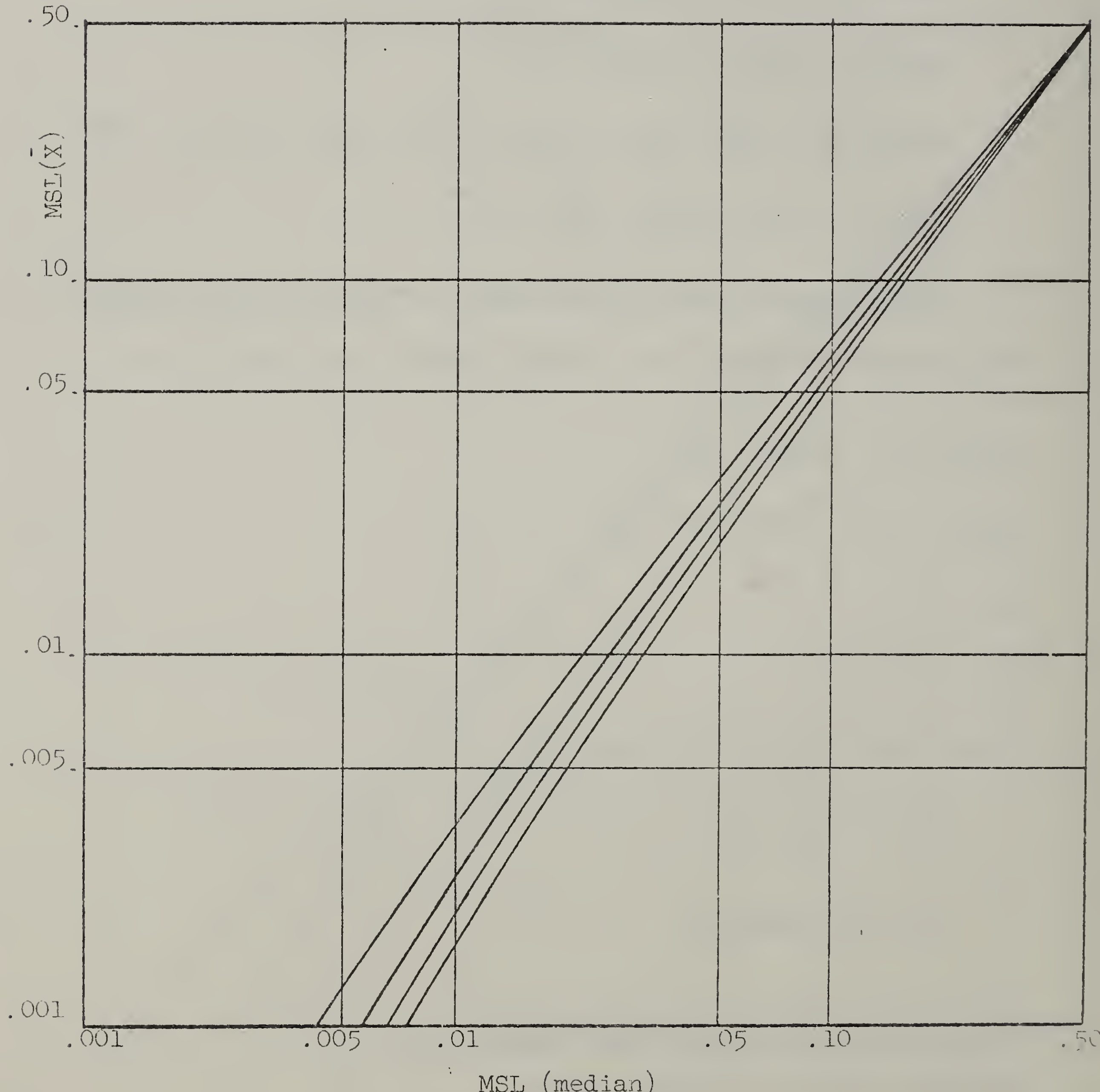
---

* Table look-up in Harter and Clemm (1959).

** Table look-up in Harter (1963).

Figure B.8

MSL($\bar{X}$) vs. MSL (median) for normal distribution. From top to botton n = 3, 5, 11, ∞ .



Since both test statistics have symmetric distributions the MSL $\doteq$ SLA. The details of computations are not included here but the method may be outlined as follows:

(1) Eisenhart, et al. (1963) table of percentage points of the sample median was used to find the alternative corresponding to a given MSL (its MCV), for the test based on the median.

(2) Tables of the normal probability integral were then used to find the MSL corresponding to that alternative, for the test based on the mean.

(3) The asymptotic values for the MSL and ESL were found by using the limiting normal approximation to the distribution of the median.

The two tests may also be easily compared for large sample sizes using the ESL criterion. In this example (for large n) the ESL curve fell within plotting error of the MSL curve. The asymptotic expressions for the ESL and MSL are:

$$ESL(\bar{X}) = \Phi\left(\frac{-\theta\sqrt{n}}{1.4142}\right)$$

$$ESL(\tilde{X}) = \Phi\left(\frac{-\theta\sqrt{n}}{1.77245}\right)$$

$$MSL(\bar{X}) \doteq \Phi\left(\frac{-\theta\sqrt{n}}{1}\right)$$

$$MSL(\tilde{X}) = \Phi\left(\frac{-\theta\sqrt{n}}{1.2533}\right) .$$

Discussion

Two relatively simple means of assessing the properties of statistical tests have been introduced and some examples have been given illustrating their usage. In many cases where it is impractical to compute a substantial portion of the power function, the MSL-MCV or SLA-ACV curve may be a useful compromise. It is believed that the use of the "built-in-metric," made possible by the ESL, MSL, or SLA may provide a useful alternative way of looking at the properties of statistical tests.

The entire distribution of the observed significance level is in general unmanageable and if one chose to use the distribution of the significance level approach, rather than the traditional power function approach, one would likely be reduced to a tabulation of its percentage points. This is seen to be equivalent to tabulating the significance level for which there is power P. If P = .50 the MSL is obtained.

In the examples given, the ESL behaved in a different manner from the MSL or SLA. The ESL is known to be equivalent to an unweighted average of the power function and its interpretation is difficult.

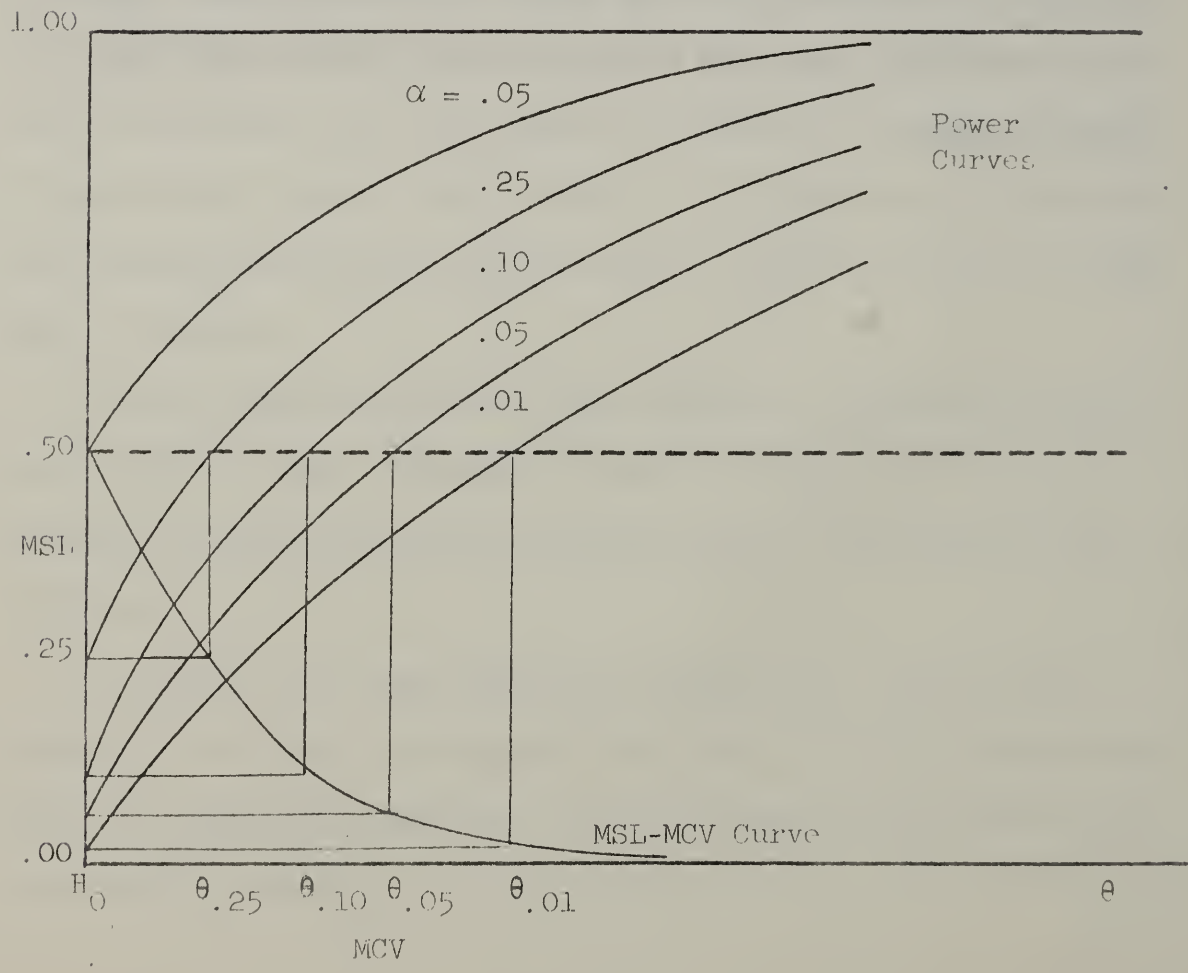
On plots of the type depicted in figures B.7 and B.8, the behavior of the ESL is more nearly like that of the MSL suggesting that for comparative purposes, the two measures would tend to give more similar results.

Figure B.8 provides an interesting look at the behavior of the MSL for two competing tests. Perhaps the most striking feature of the plot is the rapidity with which the results become "asymptotic." Put another way, in this example, asymptotic results are seen to be applicable to relatively small sample sizes.

Figure B.9 provides a way of looking at the relationships between the power, MSL and MCV functions. The MCV value for a given value of α is the θ corresponding to the intersection of the α curve and the power = .50 horizontal line. The MSL value for a given value of θ is the value of α for the power curve that passes through the intersection of the power = .50 and θ lines. The same probability scale may be used to plot the MSL curves and power as shown.

Figure B.9

Relationship between power, MSL and MCV for a hypothetical example.



The Median Significance Level and Other Small Sample Measures of Test Efficacy

Handout for IMS Meeting, Washington, D.C. 12/28/67

by Brian L. Joiner

Figure 1a

An illustration of the definitions of Geary's "average critical value" and the "median critical value." Imagine the alternative H_θ being chosen so that μ_θ or ξ_θ coincides with t_α. Then that value of θ is the ACV or MCV, respectively.

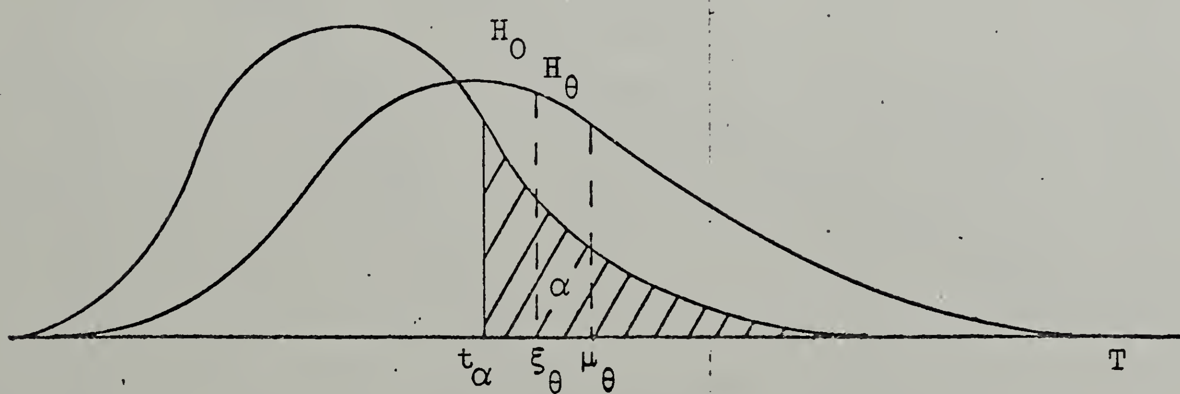


Figure 1b

A hypothetical example showing the distribution of the "observed significance level" under H_0 and under a given H_θ. The values of the "expected significance level" and the "median significance level" are also shown.

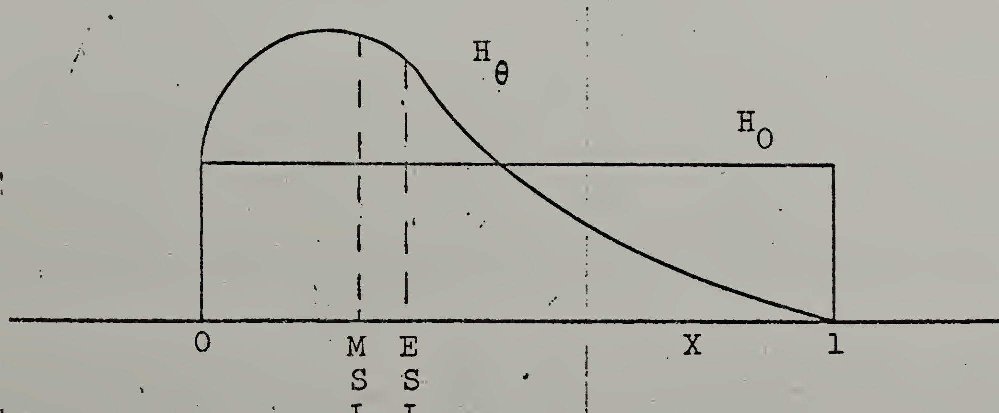


Figure 2

Upper tailed test for σ from normal distribution (n=5)

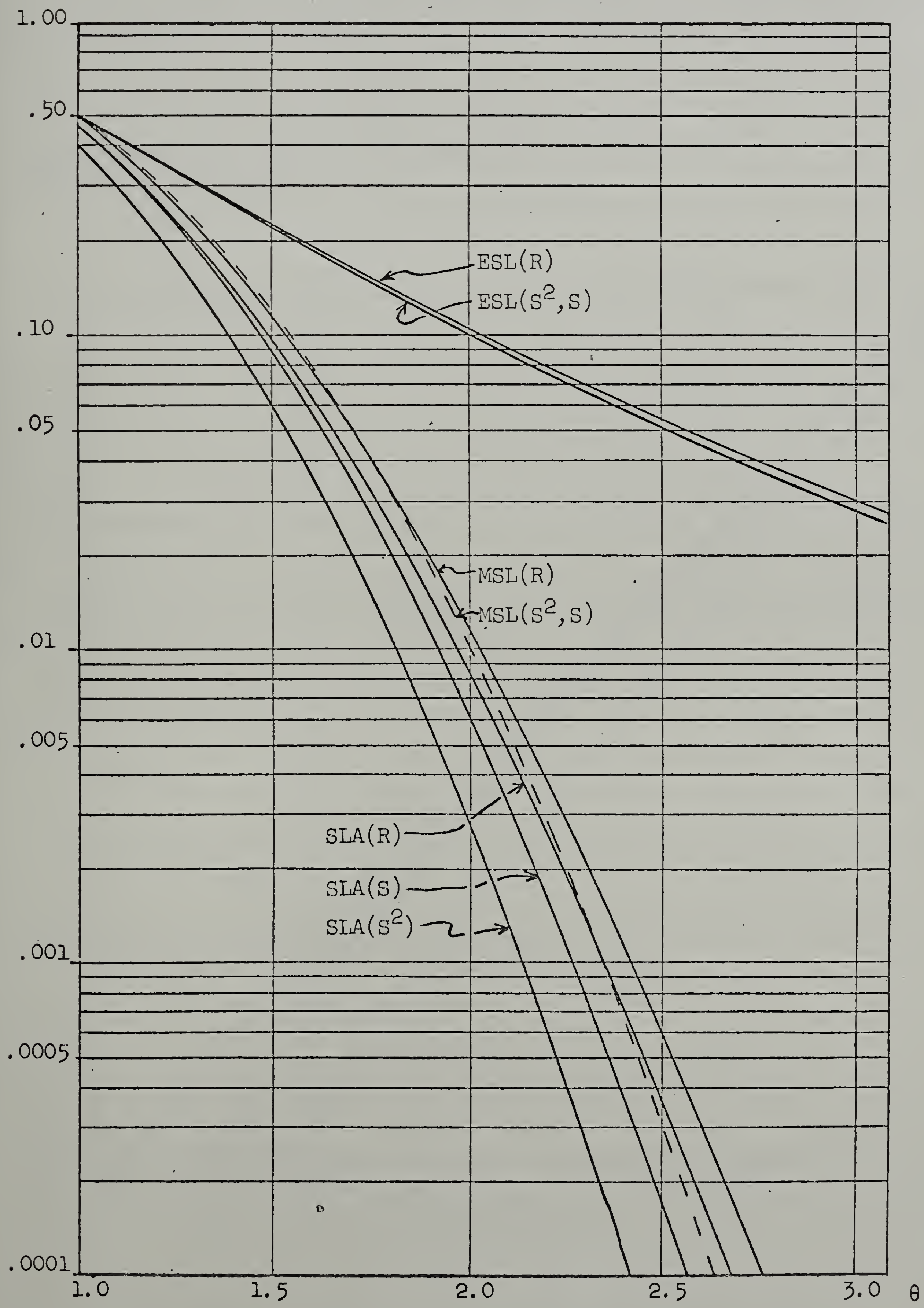


Figure 3

Upper tailed test for σ from normal distribution (n=5): a comparison of S and R.

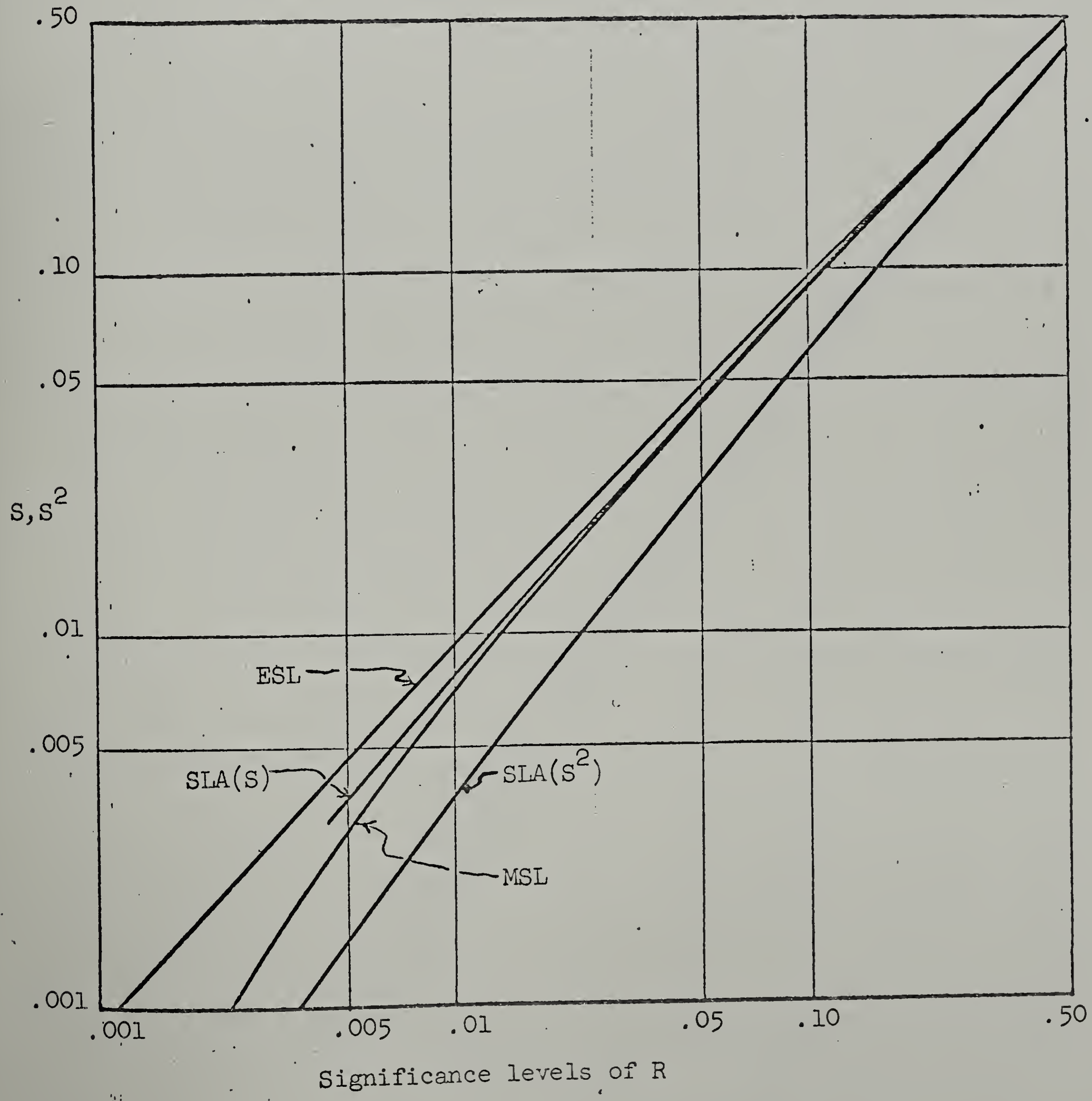
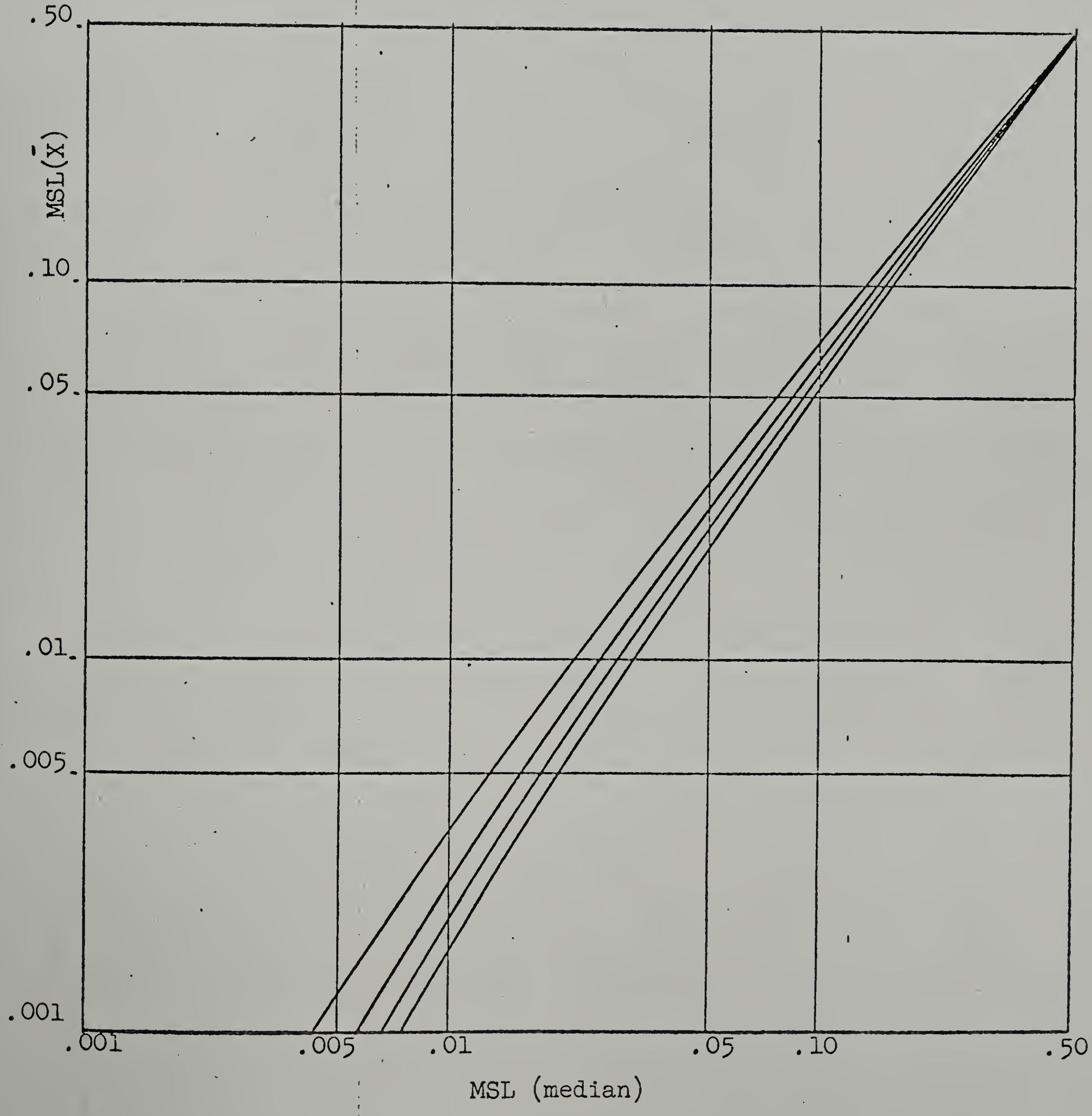


Figure 4

MSL($\bar{X}$) vs. MSL (median) for normal distribution. From top to botton n = 3, 5, 11, ∞.



DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS