

ERRATA SHEET

SCME COMPARISONS OF SENSITIVITIES FOR
FOR TWO METHODS OF MEASUREMENT

W. L. Hafley

- Page 17 - Line preceding equation (3.12) - ω_1 read ω_1 .
- Page 18 - Line preceding equation (3.14) - (3.11) read (3.12) .
Line 2 of Section 3.2.2.1 - (3.3) read (3.4) .
- Page 26 - Equation at bottom of page - (3.4) read (3.6) .
- Page 29 - Right hand side of last equation read

$$e \quad 3.29 \left(\frac{A_1}{m_1} + \frac{1}{m_2} \right)^{1/2}$$

- Page 49 - Equations (4.9) - Subscript on $1-\alpha$ read $1-\alpha$ not $1-\alpha/2$.
- Page 55 - 4th line from bottom - S' read W' .
- Page 59 - Lines 1 and 2 of 2nd paragraph - "relative sensitivity" read "relative sample size."
- Page 62 - Insert after Patnaik -
Pitman, E.J.G. 1939. A note on normal correlations. *Biometrika*, 31:9-12.

SOME COMPARISONS OF SENSITIVITIES FOR
TWO METHODS OF MEASUREMENT

by

William LeRoy Hafley

Institute of Statistics
Mimeograph Series No. 253
April, 1960

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
LIST OF FIGURES	vi
CHAPTER	
I. INTRODUCTION	1
1.1 General	1
1.2 Statement of the Problem	3
II. DISCUSSION OF LITERATURE	5
2.1 General	5
2.2 Multivariate Approach	5
2.3 Components of Measurement	6
2.4 Sensitivity Criterion	8
2.5 Comparison of Similar Experiments	11
III. THEORETICAL DEVELOPMENT	14
3.1 Introduction	14
3.2 Case I: Both Measurements Taken on the Same Experimental Units	15
3.3 Case II: Each Measurement Taken on Separate Subsamples from the Same Experimental Unit	26
3.4 Determination of Relative Sample Size	29
3.5 Power of the Tests	32
3.6 Case III: Approximation for the Schumann-Bradley Test	36
IV. APPLICATION	46
4.1 Test Procedures	46
4.2 Example of a Case I Experiment	51
4.3 Example of a Case II Experiment	53
V. SUMMARY	59
5.1 Summary	59
5.2 Suggestions for Further Research	60
LIST OF REFERENCES	61

CHAPTER I

INTRODUCTION

1.1 General

In many experimental situations the question frequently arises as to which of two or more methods of measurement should be used to observe some characteristic or response of interest. In such cases it is necessary to have a means of comparing the relative merits of the various methods. When the true value of the characteristic of interest is known, the customary procedure is to determine the accuracy of a measurement method by comparing the measured value with the true value, and to determine its standard deviation as a measure of the precision or reproducibility of the method. Accuracy and precision are then used as a basis for comparison of alternative methods. However, there are many situations in which the characteristic of interest is either ill-defined; such as, quality, performance, etc., or can only be measured by means of the instruments being compared; such as, strength, smoothness, etc. In this case it is necessary to have a means of comparing the relative merits of methods of measurement which is independent of any knowledge of the true value of the characteristic of interest.

Both of these situations, where the true value of the characteristic is either known or unknown, have been discussed by Cochran (1943) and Mandel and Stiehler (1954), and they suggested the concept of sensitivity as a qualitative measure of merit of measurement methods. Basically, the sensitivity of a measurement is a combination of its ability to detect

small variations in the characteristic of interest, or small differences between treatments; and its reproducibility. As Cochran put it:

"It is clear that the answer (to the sensitivity question) depends both on the experimental errors associated with the scales and on the magnitudes of the treatment effects in the two scales."

Thus, the most sensitive method will be that method which provides the optimum combination of high discrimination of treatment effects and high reproducibility or minimum error variance.

When the sensitivity problem is to be investigated, experiments may be carried out specifically for the purpose of comparing two methods or scales of measurement. It is assumed that it is possible to select treatments that will effect changes in the characteristic of interest, or that materials can be selected which are known to span the range of values of this characteristic; for example, see Lashof, Mandel and Worthington (1956). It is not necessary to know the true values of the characteristic, only that differences, be they large or small, exist. When at least one of the measurement processes leaves the characteristic of interest unaltered, both methods of measurement may be applied to the same experimental units. In the situation where both measurement processes alter the characteristic of interest, it is often possible to apply the two methods of measurement to separate sets of subsamples from the same experimental units. There may be situations, however, where it will be necessary to apply the methods to be compared to independent experiments, the only common feature being that the same set of treatments will be used in both experiments. This last situation has been considered by Schumann and Bradley (1957) and (1959).

Provided the techniques of analysis of variance are applicable, it is possible to obtain a comparison of sensitivities of methods of measurement by comparing the F-ratios for treatments from the analysis for each method. The most sensitive method is that one which better demonstrates treatment effects (Model I of the analysis of variance) or the existence of a between-treatments component of variance (Model II of the analysis of variance). This definition of the most sensitive measurement method conforms to that given previously.

The comparison of F-ratios is the approach considered by Schumann and Bradley. They deal with the situation where independent experiments are conducted; and, hence, the F-ratios to be compared are independent. In the situation where one experiment is conducted and both methods of measurement are applied to the same experimental units, or to subsamples of the same experimental units, the comparison of the F-ratios for treatments from the analyses of variance will involve the comparison of correlated F-ratios. It is this latter situation with which this paper is concerned.

1.2 Statement of the Problem

The primary objective of the study reported in this paper was to develop a statistical test for the comparison of the sensitivities of two methods of measurement applied to the same experiment. The approach to be considered was the comparison of the F-ratios for treatments from the analyses of variance. Two situations were of primary interest: the first being the case where both measurement methods can be applied to the same experimental units, and the second being the case where the

measurement methods must be applied to separate samples from the same experimental units. The secondary objective was to investigate the properties of the tests that were developed, particularly with regard to the power of the tests.

As an outgrowth of the study it was possible to develop an approximate test for the case considered by Schumann and Bradley (1959). They developed an exact test for the case where two independent experiments are conducted. A further development was a solution for the problem of obtaining maximum sensitivity with minimum cost. The theoretical development of these and the original objectives is presented in Chapter III. Illustrative examples of the application of the tests are given in Chapter IV.

CHAPTER II

DISCUSSION OF LITERATURE

2.1 General

The problem of determining the appropriate scale or method of measurement to be employed for the determination of a particular characteristic, or the estimation of a particular response, has been discussed throughout the literature. A great majority of the articles in which this problem has been discussed have made no attempt at presenting a general approach to the problem, but rather have dealt with the topic only in so far as it pertained to the particular study being presented. A few examples are Dillon (1936), Roth and Stiehler (1948), Carlin, Kempthorne and Gordon (1956), Lashof, et al. (1956), Teichman, et al. (1957), and Hart (1958). The only articles to be discussed in this chapter, however, are those in which an attempt has been made to present a valid statistical test for the general problem of comparing sensitivities of measurements. These latter articles are comparable to the type of study presented in this paper.

2.2 Multivariate Approach

The statistical aspects of the problem of comparing different scales of measurement for experimental results were discussed in considerable detail by Cochran (1943). He assumed that the techniques of analysis of variance were applicable and confined his attention to the case in which all scales measure the same replicated experiment. What was at that time recent work in multivariate analysis was used to provide tests of the hypothesis that the treatment effects are the same in all scales, and of the hypothesis that the scales are linearly related. All of the tests

presented were for large sample results. Aside from indicating the problems involved, no attempt was made to develop small sample tests.

A brief discussion was presented of methods for comparing the "relative sensitivity" of two scales. What Cochran refers to as relative sensitivity corresponds to what has been called the "sensitivity ratio" by other authors. Tests were suggested for comparison of two scales when only two treatments are used and the scales are either equivalent or linearly related.

For the case where more than two treatments are used it was indicated that the comparison of two scales would involve a test of the hypothesis that two non-central variance ratios are equal. However, no such test was actually developed.

2.3 Components of Measurement

Several publications have appeared in which the proposed technique for comparing two or more methods of measurement assumes that a measurement is made up of several components. In an article presented by Grubbs (1948) a measurement or observed value was considered to be the sum of two components -- one the absolute value of the characteristic measured and the other an error of measurement. He called the first component, product variability; whereas, the second component was referred to as the precision of the measurement. Techniques were given for separating and estimating product variability and precision of the measurement, and the application of these techniques to cases involving the comparison of two or more measurement methods were discussed. No significance tests were presented. The decision as to which measuring

method is superior was determined by comparing the relative order of magnitude of the estimates of the components.

Later, Smith (1950) used Grubbs' development to consider the problem of comparing two instruments for measuring a character which can only be observed via these or similar instruments; i.e., there are no theoretically absolute values against which the instruments can be calibrated. He presents a technique for computing the precision of two measuring instruments when there is a linear relation between the scales of the two instruments. A technique is presented for obtaining estimates of the components of measurement as defined by Grubbs, which requires an estimate of the regression of one scale on the other. Since Smith uses the linear relationship between the scales only as a means of obtaining estimates of the components of measurement, the outcome is independent of which regression is estimated.

Further work in the use of components was presented by Mandel (1959) and by Mandel and Lashof (1959). Both papers gave particular attention to interlaboratory studies of test methods. The assumption was made that systematic differences exist between sets of measurements made by the same observer at different times or on different instruments, or by different observers in the same or different laboratories, and that these systematic differences are linear functions of the magnitude of the measurements.

The first paper by Mandel proposes a theoretical framework for the mathematical expression of the sources of variation in measuring methods and develops a suitable method of statistical analysis. The latter paper deals with the practical aspects of the scheme, which is called

"the linear model", and demonstrates its use in the design and analysis for round-robin tests. A further discussion of the use of interlaboratory studies for the purpose of comparing methods of measurement was also presented.

2.4 Sensitivity Criterion

An alternative approach to the preceding techniques falls into a category of its own, which will be referred to as the sensitivity criterion. This criterion was presented by Mandel and Stiehler (1954). They suggested as a measure of the sensitivity of a test

$$\Psi_M = \left| \frac{\Delta M}{\Delta Q} \right| / \sigma_M, \quad (2.1)$$

where M is an obtainable measure of some property of interest, Q, and σ_M is its standard deviation. It was pointed out that the advantage of this criterion is that it takes into account, not only the reproducibility of the testing procedure, but also its ability to detect small variations in the characteristic to be measured.

For the case where we wish to compare the sensitivities of two alternative methods of measurement applied to the same experiment, the sensitivity ratio was presented in the form

$$\frac{\Psi_M}{\Psi_N} = \left| \frac{\Delta M / \Delta Q}{\Delta N / \Delta Q} \right| \cdot \frac{\sigma_N}{\sigma_M}, \quad (2.2)$$

where M and N are the alternative measuring methods. This sensitivity ratio was then reduced to

$$\frac{\Psi_M}{\Psi_N} = \left| \frac{\Delta M}{\Delta N} \right| \cdot \frac{\sigma_N}{\sigma_M}. \quad (2.3)$$

Hence, it was pointed out that it is not necessary to have a knowledge of the relation of M and N to the theoretical Q. All that is required is a knowledge of the mutual relationship between M and N. Mandel and Stiehler indicated that $\Delta M/\Delta N$ will be the slope of the curve of M plotted as a function of N. The superior method is determined by observing in which direction the sensitivity ratio deviates from one. When there is no difference in the sensitivities this ratio will equal one.

Mandel and Stiehler further stated that the functional relationship assumed to exist between the methods M and N need not be known for the application of the sensitivity ratio approach. They implied that it is only necessary to obtain the regression of M on N, and then $\Delta M/\Delta N$ will be estimated by the slope of this regression.

Let us investigate this point. If $f_c(M|N)$ denotes the conditional distribution of M given N, the regression of M on N is defined as

$$E(M|N) = \int_{-\infty}^{\infty} M [f_c(M|N)] dM, \quad (2.4)$$

and, similarly, the regression of N on M is

$$E(N|M) = \int_{-\infty}^{\infty} N [f_c(N|M)] dN, \quad (2.5)$$

Their respective slopes will therefore be

$$\frac{\Delta M}{\Delta N} = \int_{-\infty}^{\infty} M \left(\frac{\partial [f_c(M|N)]}{\partial N} \right) dM, \quad (2.6)$$

and

$$\frac{\Delta N}{\Delta M} = \int_{-\infty}^{\infty} N \left(\frac{\partial [f_c(N|M)]}{\partial M} \right) dN. \quad (2.7)$$

Now,

$$\frac{\Psi_M}{\Psi_N} = \frac{1}{\frac{\Psi_N}{\Psi_M}} \text{ implies } \left| \frac{\Delta M}{\Delta N} \right| \cdot \frac{\sigma_N}{\sigma_M} = \frac{1}{\left| \frac{\Delta N}{\Delta M} \right| \cdot \frac{\sigma_M}{\sigma_N}} \quad (2.8)$$

which implies

$$\left| \int_{-\infty}^{\infty} M \left(\frac{\partial [f_c(M|N)]}{\partial N} \right) dM \right| = 1 / \left| \int_{-\infty}^{\infty} N \left(\frac{\partial [f_c(N|M)]}{\partial M} \right) dN \right|. \quad (2.9)$$

It is well known that this last relationship is not necessarily true.

For example, if M and N come from a bivariate normal population,

$$E(M|N) = \mu_M + \rho \frac{\sigma_M}{\sigma_N} (N - \mu_N), \quad (2.10)$$

and, hence

$$\frac{\Delta M}{\Delta N} = \frac{\partial E(M|N)}{\partial N} = \rho \frac{\sigma_M}{\sigma_N}. \quad (2.11)$$

By the same token

$$\frac{\Delta N}{\Delta M} = \frac{\partial E(N|M)}{\partial M} = \rho \frac{\sigma_N}{\sigma_M}. \quad (2.12)$$

Therefore, substituting (2.11) and (2.12) in (2.8) gives

$$|\rho| = |1/\rho|, \quad (2.13)$$

which is a contradiction except when $\rho = 1$. Hence, the determination of the most sensitive method of measurement, using the slope of the regression of one method on the other as suggested by Mandel and Stiehler, is dependent on which regression is estimated.

A proposal for obtaining confidence limits for the sensitivity ratio by use of the relation

$$\left| \frac{\Delta M}{\Delta N} \right| \cdot \frac{\sigma_N}{\sigma_M} = \left| \frac{\Delta M}{\Delta N} \right| \cdot \frac{s_N}{s_M} \cdot \left(\frac{1}{F} \right), \quad (2.14)$$

where F is distributed as Snedecor's F and $\Delta M/\Delta N$ is assumed to be determined without error, was presented. This relation may be incorrect since the variances which appear in the ratio under consideration are correlated when both methods are used on the same experiment. In this case, the relation that should be used for confidence bounds on $|\Delta M/\Delta N|(\sigma_N/\sigma_M)$ is that given by Pitman (1939), namely:

$$F = \frac{\left(\frac{\Delta M}{\Delta N}\right)^2 \left(\frac{s_N}{s_M} - \frac{\sigma_N}{\sigma_M}\right)^2 (n-2)}{4(1-r_{MN}^2) \left(\frac{\Delta M}{\Delta N}\right)^2 \left(\frac{s_N \sigma_N}{s_M \sigma_M}\right)}, \quad (2.15)$$

where F is distributed as Snedecor's F with $(1, n-2)$ degrees of freedom, $r_{MN}^2 = s_{MN}^2/s_M s_N$, n is the number of samples measured, and s_N , s_M and s_{MN} are the sample standard deviations and covariances, respectively.

Aside from these points, if the functional relationship between the two methods of measurement is known, or the quality Q can be measured without error, the sensitivity criterion appears to be a satisfactory approach.

2.5 Comparison of Similar Experiments

The comparison of variance ratios from independent but similar experiments as an approach to comparing sensitivities was introduced by Schumann and Bradley (1957). They suggest that in order to compare two methods of measurement, two independent but similar experiments should be set up, and that the two scales of measurement be applied to separate experiments. The term "similar experiments" was used in the sense that the F -ratios for treatment comparisons resulting from the two experiments have the same degrees of freedom. The objective was to select that measuring method

which has the larger F-ratio, with the assumption that this would be an indication that that method was more sensitive to differences in the treatments than the method yielding the smaller F-ratio. The model assumed for the experiments was Model I, the fixed effects model of the analysis of variance as defined by Eisenhart (1947), and a test was developed for testing the hypothesis that the non-centrality parameters of the two non-central variance ratios are equal.

The distribution of the ratio of two independent non-central variance ratios was obtained and its properties discussed. Using the result of Patnaik (1949) they showed that the ratio of two non-central variance ratios may be approximated adequately by the distribution of the ratio of two central variance ratios with appropriately adjusted degrees of freedom. A table for use in applications of the latter distribution, with experiments in which the degrees of freedom for the F-ratios are the same and the treatment means are based on the same number of replications, was given for one-sided tests at the 5% level of significance. Simultaneously with this paper a companion paper was presented by Bradley and Schumann (1959) in which applications of the similar experiment approach and use of the table were presented.

Later the application of the similar experiment approach to Model II, the random effects model, of the analysis of variance as defined by Eisenhart (1947), was presented by Bross (1959), and Schumann and Bradley (1959). Both papers extended the earlier work of Schumann and Bradley (1957) on the comparison of sensitivities of experiments based on Model I of the analysis of variance to experiments based on Model II of the

analysis of variance. In addition, the Schumann and Bradley paper presented additional tables for the distribution of the ratio of two central variance ratios with equal pairs of degrees of freedom. The new tables are for the 2.5%, 1% and 0.5% levels of significance.

CHAPTER III

THEORETICAL DEVELOPMENT

3.1 Introduction

In this chapter procedures are developed for a statistical test for comparing the sensitivities of two methods of measuring some quality of a material where both methods are applied to the same experiment. Two cases are of primary interest. These are:

Case I: Both methods of measurement are applied to the same experimental units,

Case II: Each method of measurement is applied to separate subsamples from the same experimental units.

A procedure is then presented for determining the relative sample size of two measurement methods. From the relative sample size it is possible to determine the number of additional samples required with the least sensitive method to make the two methods equal in sensitivity. Cost considerations are introduced and a determination made as to which method is the least expensive to apply. The power of the tests developed for the two cases of interest is discussed, and power curves are presented for several values of the parameters. Finally, an approximation for the Schumann-Bradley test for the case of independent experiments is developed, and the critical values obtained with the approximation are compared with the exact values obtained by Schumann and Bradley. This is referred to as Case III.

The following assumptions are made throughout this chapter:

- (1) The pairs of observations on each experimental unit are assumed to come from a bivariate normal parent population.
- (2) The applicable model is Model II, the random effects model, of the analysis of variance.
- (3) The assumptions required for the use of analysis of variance procedures to infer the existence of components of variance, as described by Eisenhart (1947), are applicable to the experimental situations under consideration.

3.2 Case I: Both Measurements Taken on the Same Experimental Units

3.2.1 Notation and hypothesis. Suppose for simplicity samples are tested from m grades of a material, by two methods or for two properties, X_a and X_b . Let n_r be the number of samples from the r^{th} grade, and let $\sum_{r=1}^m (n_r) = N$. Then, from the analysis of variance and covariance for the two methods, we obtain the sums of squares and sums of products presented in Table 3.1.

Table 3.1. Analysis of variance and covariance

Source	d.f.	Sx_a^2	$Sx_a x_b$	Sx_b^2
Between Grades	m_1	a_1	c_1	b_1
Within Grades	m_2	a_2	c_2	b_2

$$m_1 = m-1, \text{ and } m_2 = \sum_{r=1}^m (n_r - 1) = N - m$$

The necessary expectations are:

$$E\left(\frac{a_1}{m_1}\right) = \sigma_w^2 + k_o \sigma_b^2, \quad E\left(\frac{b_1}{m_1}\right) = \phi_w^2 + k_o \phi_b^2, \quad (3.1)$$

$$E\left(\frac{a_2}{m_2}\right) = \sigma_w^2, \quad E\left(\frac{b_2}{m_2}\right) = \phi_w^2, \quad (3.2)$$

where k_o will equal n_r in the particular case where the same number of samples are observed in each grade.

Now, as stated in Chapter I, the goal of such an experiment is to determine which method does the best job of detecting the existence of a between-grades component of variance. Hence, the hypothesis to be tested is:

$$H_o: \frac{\sigma_b^2}{\sigma_w^2} = \frac{\phi_b^2}{\phi_w^2}, \quad (3.3)$$

against the alternative

$$H_a: \frac{\sigma_b^2}{\sigma_w^2} \neq \frac{\phi_b^2}{\phi_w^2}. \quad (3.4)$$

This alternative will be one-sided when comparing a new procedure with a standard or accepted procedure. Since the k_o 's are the same, H_o is equivalent to

$$H_o: \frac{k_o \sigma_b^2 + \sigma_w^2}{\sigma_w^2} = \frac{k_o \phi_b^2 + \phi_w^2}{\phi_w^2}, \quad (3.5)$$

which can further be written

$$H_o: \frac{k_o \frac{\sigma_b^2}{\sigma_w^2} + 1}{k_o \frac{\phi_b^2}{\phi_w^2} + 1} = \frac{\sigma_w^2}{\phi_w^2}, \quad (3.6)$$

or

$$H_o: \frac{k_o \frac{\sigma_b^2}{\sigma_w^2} + 1}{k_o \frac{\phi_b^2}{\phi_w^2} + 1} \bigg/ \frac{\sigma_w^2}{\phi_w^2} = 1. \quad (3.7)$$

Obviously, for either of the forms, the appropriate statistic is $\frac{a_1 b_2}{b_1 a_2}$.

Except for the correlation between X_a and X_b , this statistic could be expressed as a ratio of Snedecor's "F" or a difference of Fisher's "z". Since the correlation can not be ignored, the test of significance will be dependent on the c's as well as the a's and b's.

3.2.2 Adjustment for the correlation. The "between-grades" and "within-grades" lines of the analysis being independent, it is possible, from Pitman (1939), to find the distribution of the ratio of mean squares for each line, and then of their ratio. Thus, for either line of the analysis we have from Pitman

$$t_i^2 = \frac{(a_i/b_i - \omega_i)^2(m_i-1)}{4(1-r_i^2)(a_i/b_i) \omega_i}, \quad (i = 1,2); \quad (3.8)$$

where t_i is distributed as Student's "t" with (m_i-1) degrees of freedom, $r_i^2 = c_i^2/a_i b_i$, and the ω_i 's are:

$$\omega_1 = \frac{k_o \sigma_b^2 + \sigma_w^2}{k_o \phi_b^2 + \phi_w^2}, \quad (3.9)$$

and

$$\omega_2 = \frac{\sigma_w^2}{\phi_w^2}. \quad (3.10)$$

The null hypothesis may now be written

$$H_0: \omega_1 = \omega_2. \quad (3.11)$$

Now, to obtain the distribution of $\frac{a_1 b_2}{b_1 a_2}$ we must first solve (3.8) for ω_1 , which yields

$$\omega_1 = \frac{a_i}{b_i} \left\{ 1 + \frac{2t_i^2(1-r_i^2)}{(m_i-1)} \pm \frac{2t_i(1-r_i^2)^{1/2}}{(m_i-1)^{1/2}} \left[1 + \frac{t_i^2(1-r_i^2)}{(m_i-1)} \right]^{1/2} \right\}. \quad (3.12)$$

Next, let

$$d_i = \frac{t_i(1-r_i^2)^{1/2}}{(m_i-1)^{1/2}}, \quad (3.13)$$

then (3.11) can be written

$$\omega_i \left(\frac{b_i}{a_i} \right) = 1 + 2d_i^2 \pm 2d_i(1 + d_i^2)^{1/2}. \quad (3.14)$$

It is important to note that these limits are reciprocal; i.e.

$$1 + 2d_i^2 + 2d_i(1 + d_i^2)^{1/2} = 1 / [1 + 2d_i^2 - 2d_i(1 + d_i^2)^{1/2}]. \quad (3.15)$$

Finally, let

$$F_1 = 1 + 2d_1^2 + 2d_1(1 + d_1^2)^{1/2} = [d_1 + (1 + d_1^2)^{1/2}]^2, \quad (3.16)$$

and

$$F_2 = 1 + 2d_2^2 + 2d_2(1 + d_2^2)^{1/2} = [d_2 + (1 + d_2^2)^{1/2}]^2, \quad (3.17)$$

then

$$\frac{a_1 b_2}{b_1 a_2} \text{ is distributed as } \frac{F_2}{F_1} \left(\frac{\omega_1}{\omega_2} \right); \quad (3.18)$$

and, if we let $W = F_2/F_1$,

$$\frac{a_1 b_2}{b_1 a_2} \text{ is distributed as } W, \quad (3.19)$$

under $H_0^!$.

3.2.2.1 Tests of hypothesis. An α -level test of $H_0^!$ against the two-sided alternative (3.3) will be, "Reject $H_0^!$ if $W \leq W_{\alpha/2}$ or $W \geq W_{1-\alpha/2}$, otherwise do not reject $H_0^!$." To determine which method is most sensitive we need only observe that if $W \leq W_{\alpha/2}$, the method associated with X_b is the more sensitive method; if $W \geq W_{1-\alpha/2}$, the

method associated with X_a is the more sensitive method. If we fail to reject H_0 the results of the investigation indicate that there is no difference in sensitivity between methods X_a and X_b .

An alternative approach to the test is to look at the two ratios (a_1/a_2) and (b_1/b_2) and to place the larger value in the numerator of W . Hence, for the two-sided alternative, the test of significance will then be, "Reject H_0 if $W \geq W_{1-\alpha/2}$, otherwise do not reject H_0 ." In this case, provided H_0 is rejected by the test, the most sensitive method will be that method associated with the numerator of W .

3.2.2.2. Confidence bounds. Confidence bounds may be obtained for the sensitivity ratio, ω_1/ω_2 , if we first note that from the relation in (3.15) $W_{\alpha/2} = 1/W_{1-\alpha/2}$. Thus

$$\text{Prob.} \left[\frac{a_1 b_2}{b_1 a_2} / W_{1-\alpha/2} \leq \frac{\omega_1}{\omega_2} \leq \frac{a_1 b_2}{b_1 a_2} \cdot W_{1-\alpha/2} \right] = 1 - \alpha. \quad (3.20)$$

3.2.3 Distribution of the test statistic. In order to perform the statistical test described in the previous section we must first obtain the distribution of W . Remembering that $W = F_2/F_1$, we may set up the following relation:

$$\begin{aligned} u &= (1/2) \ln(F_2/F_1) , \\ &= \ln[d_2 + (1 + d_2^2)^{1/2}] - \ln[d_1 + (1 + d_1^2)^{1/2}] , \quad (3.21) \\ &= \sinh^{-1} d_2 - \sinh^{-1} d_1 . \end{aligned}$$

Next, let $d_1 = \sinh z_1$ and $d_2 = \sinh z_2$, then

$$u = z_2 - z_1 , \quad (3.22)$$

where, since they are related to the "between-grades" and "within-grades" lines of the analysis, respectively, z_1 and z_2 are independent.

From (3.13) we see that d_1 is merely a transformation on Student's t-distribution; hence, z_1 is also a transformation on Student's t-distribution; i.e.,

$$\sinh z_1 = \left(\frac{1-r_1^2}{m_1-1} \right)^{1/2} t_1, \quad (3.23)$$

where t_1 is distributed with (m_1-1) degrees of freedom. Thus,

$$\begin{aligned} f(z_1) &= \frac{1}{(1-r_1^2)^{1/2}} \frac{\Gamma\left(\frac{m_1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{m_1-1}{2}\right)} \frac{\cosh z_1}{\left(1 + \frac{\sinh^2 z_1}{1-r_1^2}\right)^{m_1/2}}, \quad -\infty < z_1 < \infty, \\ &= \frac{\Gamma\left(\frac{m_1}{2}\right)}{\Gamma\left(\frac{m_1-1}{2}\right)\Gamma\left(\frac{1}{2}\right)} (1-r_1^2)^{(m_1-1)/2} \frac{\cosh z_1}{(\cosh^2 z_1 - r_1^2)^{m_1/2}} \end{aligned} \quad (3.24)$$

Since z_1 and z_2 are independent, their joint distribution is:

$$\begin{aligned} g(z_1, z_2) &= \frac{\Gamma\left(\frac{m_1}{2}\right)\Gamma\left(\frac{m_2}{2}\right)}{\Gamma\left(\frac{m_1-1}{2}\right)\Gamma\left(\frac{m_2-1}{2}\right)} \frac{(1-r_1^2)^{(m_1-1)/2} (1-r_2^2)^{(m_2-1)/2}}{\Gamma^2\left(\frac{1}{2}\right)} \\ &\quad \times \frac{\cosh z_1 \cosh z_2}{(\cosh^2 z_1 - r_1^2)^{m_1/2} (\cosh^2 z_2 - r_2^2)^{m_2/2}}. \end{aligned} \quad (3.25)$$

Now, put $p = (z_1 + z_2)/2$ and $q = (z_2 - z_1)/2$, and substitute in (3.25) to get

$$\begin{aligned} g(p, q) &= \frac{\Gamma\left(\frac{m_1}{2}\right)\Gamma\left(\frac{m_2}{2}\right)}{\Gamma\left(\frac{m_1-1}{2}\right)\Gamma\left(\frac{m_2-1}{2}\right)} \frac{(1-r_1^2)^{(m_1-1)/2} (1-r_2^2)^{(m_2-1)/2}}{\Gamma^2\left(\frac{1}{2}\right)} \\ &\quad \frac{2 \cosh(p-q) \cosh(p+q)}{[\cosh^2(p-q) - r_1^2]^{m_1/2} [\cosh^2(p+q) - r_2^2]^{m_2/2}}. \end{aligned} \quad (3.26)$$

Thence, by integrating p out of this expression, we can obtain the distribution of q ; and, since $u = 2q$, we would then be able to obtain the distribution of u .

As an alternative approach, calculate the characteristic function of z_i . Then, if we note that $\cosh z = \cosh(-z)$, the characteristic function of u is the product of the characteristic functions of the z_i 's. Thus, the inversion procedures presented by Cramér (1957 sec. 10.3) could be applied in order to obtain the distribution of u . Let it suffice to say that this approach requires an equally intricate integration to that required for the approach using equation (3.26).

The characteristic function of z_i , $\phi_{z_i}(\theta)$, will be obtained here, however. By rewriting equation (3.24) in the form

$$f(z_i) = K(\cosh z_i)^{1-m_i} \left(1 - \frac{r_i^2}{\cosh^2 z_i}\right)^{-m_i/2}, \quad (3.27)$$

and expanding the negative binomial, we get

$$\begin{aligned} f(z_i) &= K(\cosh z_i)^{1-m_i} \left[1 + \frac{m_i}{2} \left(\frac{r_i}{\cosh z_i}\right)^2 + \frac{\frac{m_i}{2}(\frac{m_i}{2}+1)}{2!} \left(\frac{r_i}{\cosh z_i}\right)^4 + \dots \right], \\ &= K \sum_{s=0}^{\infty} \frac{\frac{m_i}{2}(\frac{m_i}{2}+1) \dots (\frac{m_i}{2}+s-1)}{s!} r_i^{2s} (\cosh z_i)^{1-m_i-2s}. \end{aligned} \quad (3.28)$$

Thus,

$$\begin{aligned} \phi_{z_i}(\theta) &= K \sum_{s=0}^{\infty} \frac{\Gamma(\frac{m_i}{2}+s) r_i^{2s}}{\Gamma(\frac{m_i}{2}) \Gamma(s+1)} \int_{-\infty}^{\infty} \frac{e^{i\theta z_i}}{(e^{z_i} + e^{-z_i})^{2s+m_i-1}} dz_i, \\ &= \frac{(1-r_i^2)^{(m_i-1)/2}}{\Gamma(\frac{m_i-1}{2}) \Gamma(\frac{1}{2})} \sum_{s=0}^{\infty} \frac{\Gamma(\frac{m_i}{2}+s) r_i^{2s}}{\Gamma(s+1)} \frac{\Gamma(\frac{m_i+2s-1}{2} + \frac{i\theta}{2}) \Gamma(\frac{m_i+2s-1}{2} - \frac{i\theta}{2})}{2 \Gamma(m_i+2s-1)}. \end{aligned} \quad (3.29)$$

Equation (3.29) may then be used to generate the moments of z_i . However, there is a more direct approach. From (3.23) we have that $z_i = \sinh^{-1} \lambda_i t_i$, where $\lambda_i^2 = \frac{1-r_i^2}{m_i-1}$, and t_i follows the t -distribution with (m_i-1) degrees of freedom. Obviously then, z_i is distributed symmetrically about zero; therefore,

$$E(z_i^{2s+1}) = 0, \quad s = 0, 1, 2, \dots$$

(For the remainder of the development of the moments of z_i the subscript, i , will be omitted.) Expanding z in a Taylor's series about zero gives:

$$z = \lambda t - \frac{1}{2} \frac{(\lambda t)^3}{3} + \frac{1 \cdot 3}{2 \cdot 4} \frac{(\lambda t)^5}{5} - \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} \frac{(\lambda t)^7}{7} + \dots$$

Thus,

$$z^2 = (\lambda t)^2 - \frac{(\lambda t)^4}{3} + \frac{8(\lambda t)^6}{45} - \frac{4(\lambda t)^8}{35} + \dots,$$

$$z^4 = (\lambda t)^4 - \frac{2(\lambda t)^6}{3} + \frac{7(\lambda t)^8}{15} - \dots,$$

$$z^6 = (\lambda t)^6 - (\lambda t)^8 + \dots,$$

$$z^8 = (\lambda t)^8 - \dots,$$

and from the moments of the t -distribution [Cramér (1957)],

$$\begin{aligned} E(z^2) &= \left(\frac{1-r^2}{m-1}\right) \frac{(m-1)}{(m-3)} - \frac{1}{3} \left(\frac{1-r^2}{m-1}\right)^2 \frac{3(m-1)^2}{(m-3)(m-5)} + \frac{8}{45} \left(\frac{1-r^2}{m-1}\right)^3 \frac{15(m-1)^3}{(m-3)(m-5)(m-7)} \\ &= \frac{4}{35} \left(\frac{1-r^2}{m-1}\right)^4 \frac{105(m-1)^4}{(m-3)(m-5)(m-7)(m-9)} + \dots \\ &= \frac{1-r^2}{m-3} - \frac{(1-r^2)^2}{(m-3)(m-5)} + \frac{8(1-r^2)^3}{3(m-3)(m-5)(m-7)} - \frac{12(1-r^2)^4}{(m-3)(m-5)(m-7)(m-9)} + \dots \end{aligned}$$

Similarly

$$E(z^4) = \frac{3(1-r^2)^2}{(m-3)(m-5)} - \frac{10(1-r^2)^3}{(m-3)(m-5)(m-7)} + \frac{49(1-r^2)^4}{(m-3)(m-5)(m-7)(m-9)} - \dots,$$

$$E(z^6) = \frac{15(1-r^2)^3}{(m-3)(m-5)(m-7)} - \frac{105(1-r^2)^4}{(m-3)(m-5)(m-7)(m-9)} + \dots,$$

and

$$E(z^8) = \frac{105(1-r^2)^4}{(m-3)(m-5)(m-7)(m-9)} - \dots.$$

These moments may now be simplified by expanding the denominators in a binomial series, as follows:

$$(m-3)^{-1} = m^{-1} + 3m^{-2} + 9m^{-3} + 27m^{-4} + \dots,$$

$$(m-3)^{-1}(m-5)^{-1} = m^{-2} + 8m^{-3} + 49m^{-4} + \dots,$$

$$(m-3)^{-1}(m-5)^{-1}(m-7)^{-1} = m^{-3} + 15m^{-4} + \dots,$$

$$(m-3)^{-1}(m-5)^{-1}(m-7)^{-1}(m-9)^{-1} = m^{-4} + \dots.$$

Letting $A = 1-r^2$, the moments of z can be written as:

$$E(z^2) = \frac{A}{m} + \frac{3A-A^2}{m^2} + \frac{9A-8A^2 + 8/3A^3}{m^3} + \frac{27A-49A^2 + 40A^3-12A^4}{m^4} + \dots,$$

$$E(z^4) = \frac{3A^2}{m^2} + \frac{24A^2-10A^3}{m^3} + \frac{147A^2 - 150A^3 + 49A^4}{m^4} + \dots,$$

$$E(z^6) = \frac{15A^3}{m^3} + \frac{225A^3-105A^4}{m^4} + \dots,$$

$$E(z^8) = \frac{105A^4}{m^4} + \dots.$$

Due to the independence of z_1 and z_2 the moments of u may be obtained by taking the expectation of the expansion of $u^n = (z_2 - z_1)^n$, $n = 1, 2, \dots$. Since z_1 and z_2 are symmetrically distributed about zero, u is also symmetrically distributed about zero. Hence,

$$E(u^{2s+1}) = 0, \quad s = 0, 1, 2, \dots,$$

and the even moments are:

$$E(u^2) = \left(\frac{A_1}{m_1} + \frac{A_2}{m_2} \right) + \left(\frac{3A_1 - A_1^2}{m_1^2} + \frac{3A_2 - A_2^2}{m_2^2} \right) + \left(\frac{9A_1 - 8A_1^2 + 8/3 A_1^3}{m_1^3} + \frac{9A_2 - 8A_2^2 + 8/3 A_2^3}{m_2^3} \right) + \dots,$$

$$E(u^4) = \left(\frac{3A_1^2}{m_1^2} + \frac{3A_2^2}{m_2^2} \right) + \frac{6A_1 A_2}{m_1 m_2} + \left(\frac{24A_1^2 - 10A_1^3}{m_1^3} + \frac{24A_2^2 - 10A_2^3}{m_2^3} \right) + 6 \left(\frac{3A_1 A_2 - A_1 A_2^2}{m_1 m_2^2} + \frac{3A_2 A_1 - A_2 A_1^2}{m_2 m_1^2} \right) + \dots,$$

$$E(u^6) = \left(\frac{15A_1^3}{m_1^3} + \frac{15A_2^3}{m_2^3} \right) + \left(\frac{45A_1 A_2^2}{m_1 m_2^2} + \frac{45A_2 A_1^2}{m_2 m_1^2} \right) + \dots,$$

$$E(u^8) = \left(\frac{105A_1^4}{m_1^4} + \frac{105A_2^4}{m_2^4} \right) + \dots.$$

Since m_1 will in general be larger than m_2 , it can be seen by inspection, and by noting that $0 \leq A_i \leq 1$, that all terms of third order or higher in m_1^{-1} will be negligible, even for values of m_1 as small as ten. Thus, dropping terms of third order and higher, the moments of u become

$$\sigma_u^2 = \frac{A_1}{m_1} + \frac{A_2}{m_2} + \frac{3A_1 - A_1^2}{m_1^2} + \frac{3A_2 - A_2^2}{m_2^2}, \quad (3.30)$$

$$\mu_4(u) = 3 \left(\frac{A_1}{m_1} + \frac{A_2}{m_2} \right)^2. \quad (3.31)$$

Now, since u is symmetrically distributed about zero, the skewness coefficient of u is zero; and, if the third order terms are omitted,

the kurtosis coefficient of u is three. These are the skewness and kurtosis coefficients of the normal distribution, thus for values of m_1 and m_2 for which it is reasonable to drop the higher order terms in m_1^{-1} , u/σ_u approximates the standard normal distribution.

For values of m_1 between five and ten, the third order terms of the moments must be retained; thus,

$$\sigma_u^2 = \frac{A_1}{m_1} + \frac{A_2}{m_2} + \frac{3A_1 - A_1^2}{m_1^2} + \frac{3A_2 - A_2^2}{m_2^2} + \frac{9A_1 - 8A_1^2 + 8/3 A_1^3}{m_1^3} + \frac{9A_2 - 8A_2^2 + 8/3 A_2^3}{m_2^3}; \quad (3.32)$$

while, for values of m_1 equal to or greater than thirty-five, second and third order terms in m_1^{-1} of σ_u^2 may be dropped, giving

$$\sigma_u^2 = \frac{A_1}{m_1} + \frac{A_2}{m_2}. \quad (3.33)$$

When m_1 is less than five the normal approximation should not be used. Note that the higher the correlation between X_a and X_b the smaller the values of A_1 and A_2 will be; and, hence, the better the normal approximation will fit.

Since the objective of this section was to obtain the distribution of W , note that $u = (1/2) \ln(W)$ and that, as has been shown above, u is approximately normally distributed with mean zero and variance σ_u^2 . Then W has approximately a lognormal distribution with mean $e^{2\sigma_u^2}$ and variance $e^{4\sigma_u^2} (e^{4\sigma_u^2} - 1)$.

If $\xi = u/\sigma_u$, then to calculate the critical value $W_{1-\alpha/2}$ we need first obtain the critical value $\xi_{1-\alpha/2}$ corresponding to the area under the normal curve of $1-\alpha/2$. Thus, $W_{1-\alpha/2} = e^{2\sigma_u^2} \xi_{1-\alpha/2}^2$. For example,

if $\alpha = .05$ and m_1 is large,

$$W_{.975} = e \quad 3.29 \left(\frac{A_1}{m_1} + \frac{A_2}{m_2} \right)^{1/2}$$

3.3 Case II: Each Measurement Taken on Separate Subsamples from the Same Experimental Unit

3.3.1 Hypothesis and notation. Suppose now that two test procedures, or methods of measurement, are being compared which are both of a destructive nature. In this case it is not possible to apply both tests to the same experimental units. The experimental design here would require that $2n_r$ samples be taken from the r^{th} grade of the material under study. These $2n_r$ samples would then be divided equally between the two test procedures so that each procedure is applied to n_r samples. If $\sum_{r=1}^m (n_r) = N$, the analysis of variance and covariance will be the same as in Section 3.2.1, Table 3.1, except that the sums of squares from the "within-grades" line of the analysis are now independent; i.e., the sum of products term c_2 will not appear.

The hypothesis to be tested is exactly the same as for Case I; i.e.,

$$H_0^1 : \frac{k_o \sigma_b^2 + \sigma_w^2}{k_o \phi_b^2 + \phi_w^2} = \frac{\sigma_w^2}{\phi_w^2} \quad (3.4)$$

However, for this case the right-hand ratio is a ratio of independent variances, and its estimate a_2/b_2 is distributed as $F_0(\sigma^2/\phi^2)$, where F_0 follows an F-distribution with m_2 degrees of freedom for both the numerator and denominator.

It should be noted that the approach for Case II is not restricted to the situation where both methods are of a destructive nature. Even though both methods of measurement may be applied to the same sample, it may be desirable to use separate samples for each method. This might occur if a time effect exists between successive measurements on the same experimental unit and we wish to eliminate it in this way. It will be shown in Section 3.5, however, that the test for Case II is not as statistically powerful as the test for Case I.

3.3.2 Distribution of the test statistic. The test of hypothesis and confidence bounds discussed in Sections 3.2.2.1 and 3.2.2.2 are also applicable to Case II; however, the test statistic

$$\frac{a_1 b_2}{b_1 a_2} \text{ is now distributed as } \frac{F_0}{F_1} \left(\frac{\omega_1}{\omega_2} \right) ; \quad (3.34)$$

where F_0 is distributed as Snedecor's F with m_2 degrees of freedom for both the numerator and denominator, and F_1 is as in Section 3.2.2. Let $W' = F_0/F_1$, then under H'

$$\frac{a_1 b_2}{b_1 a_2} \text{ is distributed as } W'. \quad (3.35)$$

Next, to obtain the distribution of W' , let

$$\begin{aligned} u' &= (1/2) \ln(F_0/F_1) , \\ &= (1/2) \ln(F_0) - (1/2) \ln(F_1) , \\ &= z_0 - z_1 , \end{aligned} \quad (3.36)$$

where z_0 is distributed as Fisher's z-distribution with (m_2, m_2) degrees of freedom, z_1 is as in equation (3.24), and z_0 and z_1 are independent.

The moments of z_1 have been presented in Section 3.2.3; thus, to obtain the moments of u' we need the moments of z_0 . First, note that when the numerator and denominator degrees of freedom are equal, Fisher's z -distribution is symmetrically distributed about zero; hence,

$$E(z_0^{2s+1}) = 0, \quad s = 0, 1, 2, \dots$$

The even moments of z_0 are [Cornish and Fisher (1937)]:

$$E(z_0^2) = \frac{1}{m_1} + \frac{1}{m_2} + \frac{2}{3m_2^3},$$

$$E(z_0^4) = \frac{3}{m_2} + \frac{8}{m_2^3} + \frac{13}{m_2^4},$$

$$E(z_0^6) = \frac{30}{m_2^4} + \frac{111}{m_2^5} + \frac{110}{m_2^6} + \frac{60}{m_2^7}.$$

Therefore, since z_0 and z_1 are independent, $u' = z_0 - z_1$ is also symmetrically distributed about zero, and its moments are

$$E(u'^{2s+1}) = 0, \quad s = 0, 1, 2, \dots$$

$$E(u'^2) = \left(\frac{A_1}{m_1} + \frac{1}{m_2} \right) + \left(\frac{3A_1 - A_1^2}{m_1^2} + \frac{1}{m_1} \right) + \left(\frac{9A_1 - 8A_1^2 + 8/3 A_1^3}{m_1^3} + \frac{2}{3m_2^3} \right) + \dots,$$

$$E(u'^4) = \left(\frac{3A_1^2}{m_1^2} + \frac{3}{m_2} \right) + \frac{6A_1}{m_1 m_2} + \left(\frac{24A_1^2 - 10A_1^3}{m_1^3} + \frac{8}{m_2^3} \right) + 6 \left(\frac{3A_1 - A_1^2}{m_1 m_2} + \frac{A_1}{m_1 m_2^2} \right) + \dots,$$

$$E(u'^6) = \frac{15A_1^3}{m_1^3} + \frac{45A_1^2}{m_1 m_2} + \frac{45A_1^2}{m_1 m_2^2} + \dots$$

Referring to the discussion at the end of Section 3.2.3 we see that it is reasonable to drop higher order terms of the moments. Thus, for m_1 greater than or equal to ten,

$$\sigma_{u'}^2 = \left(\frac{A_1}{m_1} + \frac{1}{m_2} \right) + \left(\frac{3A_1 - A_1^2}{m_1^2} + \frac{1}{m_2} \right), \quad (3.37)$$

and

$$\mu_4(u') = 3 \left(\frac{A_1}{m_1} + \frac{1}{m_2} \right)^2 \quad (3.38)$$

Again a normal approximation may be used in that $u'/\sigma_{u'}$ is approximately distributed as a standard normal. Hence, W' is approximately distributed as a lognormal with mean $e^{2\sigma_{u'}^2}$ and variance $e^{4\sigma_{u'}^2}(e^{4\sigma_{u'}^2} - 1)$.

For values of m_1 between five and ten, use

$$\sigma_{u'}^2 = \left(\frac{A_1}{m_1} + \frac{1}{m_2} \right) + \left(\frac{3A_1 - A_1^2}{m_1^2} + \frac{1}{m_2} \right) + \left(\frac{9A_1 - 8A_1^2 + 8/3 A_1^3}{m_1^3} + \frac{2}{3m_2^3} \right); \quad (3.39)$$

and, for values of m_1 greater than thirty-five, use

$$\sigma_{u'}^2 = \frac{A_1}{m_1} + \frac{1}{m_2} \quad (3.40)$$

The proposed test should again not be used for values of m_1 less than five.

To calculate the critical value $W'_{1-\alpha/2}$,

$$W'_{1-\alpha/2} = e^{2\sigma_{u'}^2 \xi_{1-\alpha/2}} \quad (3.41)$$

where $\xi_{1-\alpha/2}$ is the normal deviate corresponding to an area under the normal curve of $1-\alpha/2$. For example if $\alpha = .05$ and m_1 is large,

$$W'_{.975} = e^{3.29 \left(\frac{A_1}{m_1} + \frac{1}{m_2} \right)}$$

3.4 Determination of relative sample size. In conjunction with any comparison of sensitivities of measurement methods, we may also be interested in comparing the cost of obtaining a given sensitivity. Suppose, for example, in an acceptance-testing situation it has been

determined that we should test n samples from a particular shipment of material, using a standard test procedure. Further suppose that an experiment designed to compare sensitivities has shown this accepted method, say X_a , to be more sensitive than some alternative method, say X_b . Now, the question arises as to how many samples would be required with method X_b to attain the same sensitivity that we have using n samples and method X_a ? And, using these sample sizes, which method has the minimum cost? By referring to the hypothesis actually being tested, we can provide an answer to these questions.

To obtain an answer to this problem, let us first consider what this difference in sensitivity implies. When we say that method X_a is more sensitive than method X_b , we are saying that

$$\frac{\sigma_b^2}{\sigma_w^2} > \frac{\phi_b^2}{\phi_w^2} \quad (3.42)$$

Now, if n samples are observed in each grade, (3.42) may be written

$$\frac{n\sigma_b^2}{\sigma_w^2} > \frac{n\phi_b^2}{\phi_w^2} \quad (3.43)$$

Hence, in order to make these two sensitivities equal we must obtain a multiplier for the right-hand side of (3.43); i.e.,

$$\frac{n\sigma_b^2}{\sigma_w^2} = \frac{h_{ba} n \phi_b^2}{\phi_w^2} \quad (3.44)$$

where h_{ba} is actually the relative sample size of method X_b given the sample size to be used with method X_a . The sample size to be used with method X_b to attain a sensitivity equal to that given by method X_a when n samples are used will then be $h_{ba} n$.

The first step in obtaining the value of h_{ba} is to obtain estimates of σ_b^2 , σ_w^2 , ϕ_b^2 , and ϕ_w^2 from the analysis of variance and covariance. These will be:

$$\begin{aligned} \hat{\sigma}_b^2 &= \left(\frac{a_1}{m_1} - \frac{a_2}{m_2} \right) / k_o, & \hat{\phi}_b^2 &= \left(\frac{b_1}{m_1} - \frac{b_2}{m_2} \right) / k_o, \\ \hat{\sigma}_w^2 &= \frac{a_2}{m_2} & \text{and} & \hat{\phi}_w^2 = \frac{b_2}{m_2}. \end{aligned} \quad (3.45)$$

Thus, substituting these estimates in (3.44) and solving for h_{ba} , we get

$$h_{ba} = \frac{\hat{\sigma}_b^2 / \hat{\phi}_b^2}{\hat{\sigma}_w^2 / \hat{\phi}_w^2}. \quad (3.46)$$

To determine the total cost, if C_a is the cost per sample using method X_a , and C_b is the cost per sample using method X_b , then

$$T_a = C_a n, \quad \text{and} \quad T_b = (C_b h_{ba} n) \quad (3.47)$$

are the respective total costs. If T_b is less than T_a , method X_b is the most economical for a given sensitivity; and, if T_a is less than T_b , method X_a is the most economical.

If, instead, the relative sample size of method X_a , given the sample size to be used with method X_b , is desired, (3.44) will be written

$$\frac{h_{ab} n \sigma_b^2}{\sigma_w^2} = \frac{n \phi_b^2}{\phi_w^2}. \quad (3.48)$$

Substituting the estimates of the components of variance in (3.48) and solving for h_{ab} , we get

$$h_{ab} = \frac{\hat{\phi}_b^2 / \hat{\phi}_w^2}{\hat{\sigma}_b^2 / \hat{\sigma}_w^2}. \quad (3.49)$$

Note that $h_{ab} = 1/h_{ba}$. The total costs in this instance will be

$$T_a = (C_a h_{ab} n) \quad , \quad \text{and} \quad T_b = C_b n \quad , \quad (3.50)$$

respectively.

It must be pointed out that the development in this section is not a procedure for determining the sample size required for a particular sampling problem. It is assumed that the optimum sample size has been determined previously; and, hence, the number of samples required for at least one of the methods is known.

3.5 Power of the tests. In order to assess the value of the tests described in Sections 3.2 and 3.3, it is necessary to look at the power of the tests. The power of a test is defined as the probability of making the correct decision when, in fact, the alternative hypothesis is true. Since the main goal of the test of sensitivities is to select the best measuring procedure for a particular type of problem, it is important to know how capable the test is of detecting differences when they exist.

In both Case I and II when the normal approximation is appropriate it has been pointed out that the test statistic, W , follows the log-normal distribution. Thus, the first step in calculating the power is to obtain the critical values of the test under H_0^I .

As shown previously, the critical values under H_0^I for the two-tailed test are

$$W_{1-\alpha/2}^0 = e^{2\sigma u_{1-\alpha/2}} \quad ,$$

and

$$W_{\alpha/2}^0 = 1/W_{1-\alpha/2}^0 \quad ;$$

(3.51)

where ξ is distributed $N(0,1)$. The critical values under the alternative will be

$$W_U^a = W_{1-\alpha/2}^0 / (\omega_1/\omega_2) , \quad (3.52)$$

and

$$W_L^a = 1/(W_{1-\alpha/2}^0) (\omega_1/\omega_2) .$$

Hence, for a particular simple alternative hypothesis; i.e., a particular value of ω_1/ω_2 , the power of the test may be calculated from the relation

$$\text{Power} = \text{Prob.}(W \leq W_L^a | \omega_1/\omega_2) + \text{Prob.}(W \geq W_U^a | \omega_1/\omega_2) . \quad (3.53)$$

Remembering that $u = (1/2) \ln(W)$, (3.53) can be written

$$\text{Power} = \text{Prob.}(u \leq u_L^a | \omega_1/\omega_2) + \text{Prob.}(u \geq u_U^a | \omega_1/\omega_2) , \quad (3.54)$$

where

$$\begin{aligned} u_U^a &= (1/2) \ln(W_U^a) , \\ &= \sigma_u \xi_{1-\alpha/2} - (1/2) \ln(\omega_1/\omega_2) , \end{aligned} \quad (3.55)$$

and

$$\begin{aligned} u_L^a &= (1/2) \ln(W_L^a) , \\ &= -[\sigma_u \xi_{1-\alpha/2} + (1/2) \ln(\omega_1/\omega_2)] . \end{aligned} \quad (3.56)$$

To obtain the desired probabilities for equation (3.54) it is only necessary to consult a table of the Cumulative Normal Distribution, since u^2/σ_u is distributed $N(0,1)$.

Figure 3.1 presents power curves of the tests of sensitivity for $\alpha = .05$, $m_1 = 10$, $m_2 = 55$ and three values of r_1 and r_2 . For simplicity, r_1 and r_2 have been set equal to each other; however, this would not generally be the case. Curves for both Case I and Case II are presented on the same figure. Note that except for Figure 3.1 (a), where the two

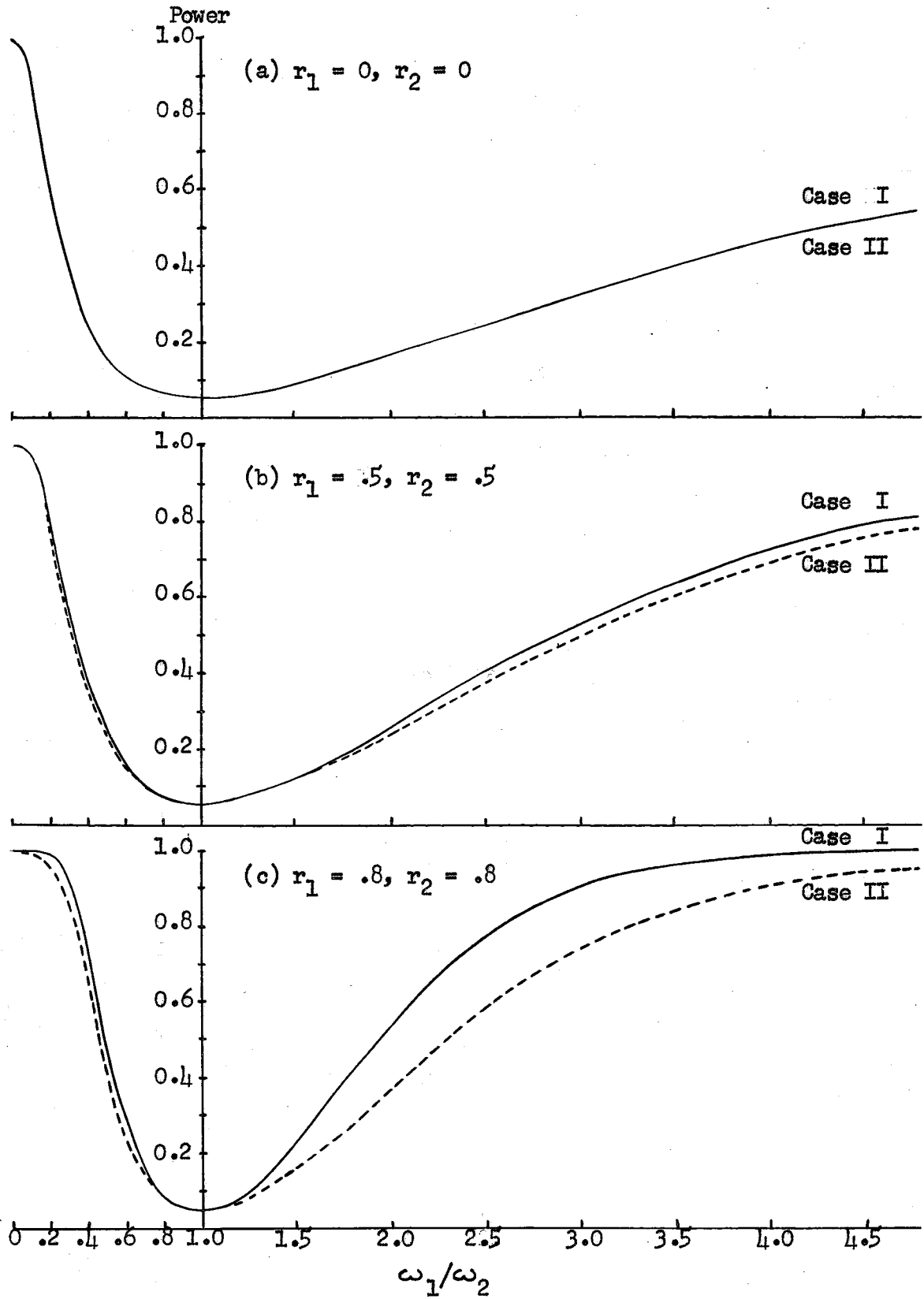


Figure 3.1. Power curves of the tests for sensitivity

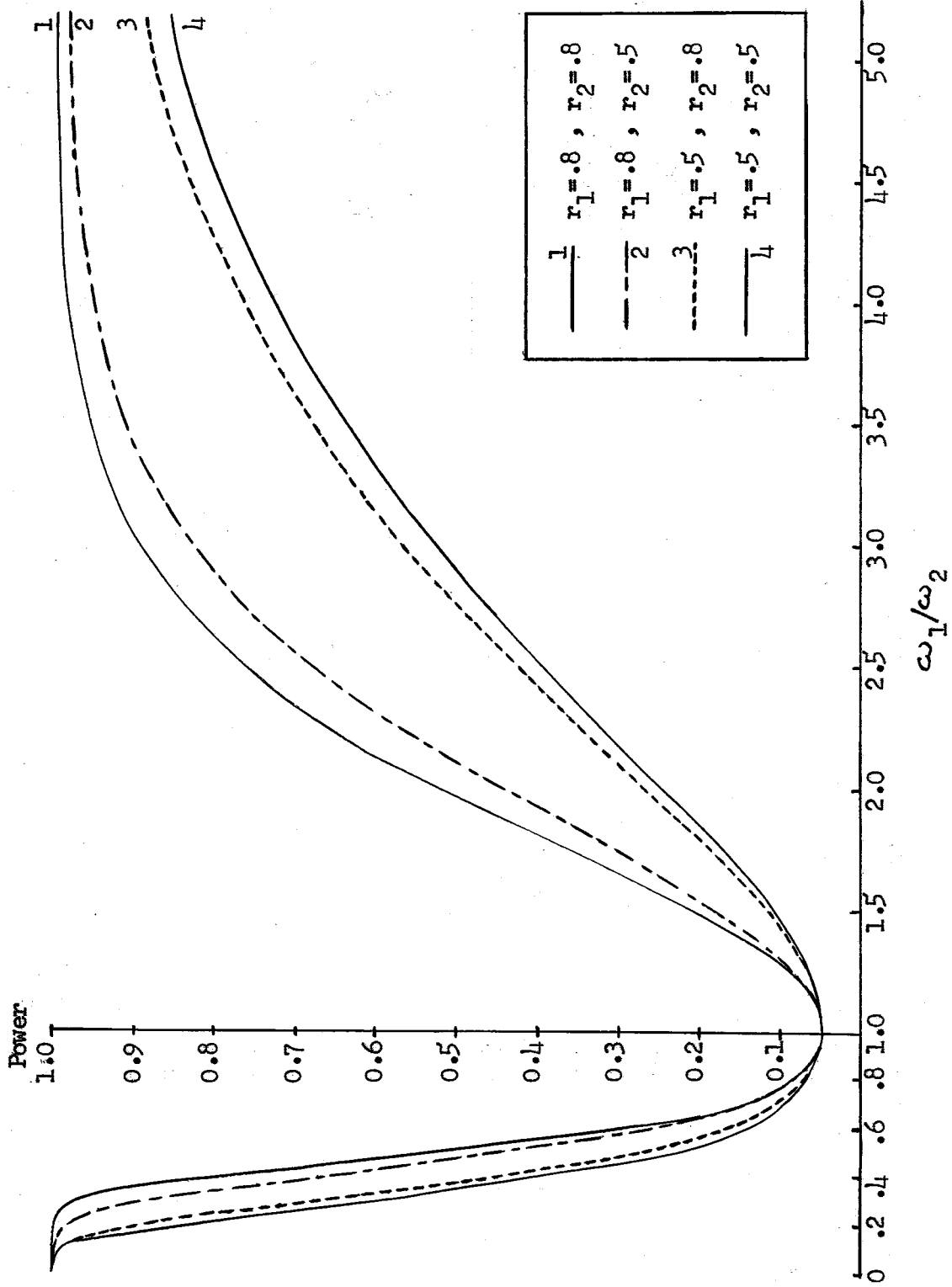


Figure 3.2. Power curves of the test for sensitivity, indicating the effects of changes in correlation

cases have the same power, the test for Case I is uniformly more powerful than the test for Case II. Note also that the power increases as the correlation between X_a and X_b increases. In fact, the power is inversely proportional to σ_u ; and, hence, an increase in the correlation or in the degrees of freedom, which will in turn bring about a decrease in σ_u , will result in an increase in power.

Figure 3.2 again presents power curves of the tests of sensitivity for $\alpha = .05$ and $m_1 = 10$, $m_2 = 55$, but with various combinations of values of r_1 and r_2 . The comparison of curve 1 with curve 2, and curve 3 with curve 4, illustrates the effect of changes in r_2 , while the comparison of curve 1 with curve 3, and curve 2 with curve 4, illustrates the effect of changes in r_1 . Note that changes in r_1 have a marked effect on power, whereas changes in r_2 have a negligible effect on power. Also, note that any increase in correlation increases the power.

It has been pointed out previously that even if both methods of measurement can be applied to the same sample, it is still valid to use an experimental design which would be classified under Case II. However, the reduction in power from Case I to Case II, illustrated in Figure 3.1, would indicate that a design suitable to Case I should be used.

3.6 Case III: Approximation for the Schumann-Bradley Test

3.6.1 Notation and hypothesis. As a means of demonstrating the closeness of the approximations developed in Sections 3.2 and 3.3 we will develop an approximation for the Schumann-Bradley test, and compare the critical values obtained by the approximation to the

exact values obtained by Schumann and Bradley (1959). The experimental situation considered by Schumann and Bradley is the case in which two independent experiments are conducted and a different method of measurement is applied to each experiment. They have developed an exact test for the comparison of the sensitivities of the two measurement methods which involves a ratio of the F-ratios for treatments from the analyses of variance for the two independent experiments. As explained previously, the use of their tables requires that the experiments be identical; i.e., that the degrees of freedom for the F-ratios for treatments and the number of observations for a given treatment are the same in both experiments. Using the approach of Sections 3.2 and 3.3, it is possible to develop an approximate test for this case.

Suppose m treatments are selected which will produce differences in the characteristic of interest. Let n_r be the number of samples observed in the r^{th} treatment, and let $\sum_{r=1}^m (n_r) = N$. Next, suppose that two experiments are conducted in which the same m treatments are used, and n_r samples are observed in the r^{th} treatment in each experiment. Further, suppose that measurement method X_a is applied to one experiment, and measurement method X_b is applied to the other experiment. Then the sums of squares from the analysis of each method would be the same as in Table 3.1. For this situation, however, the sums of products terms, c_1 and c_2 , would not appear.

The hypothesis to be tested is exactly the same as for the previous two cases; i.e.,

$$H_0^* : \frac{k_o \sigma_b^2 + \sigma_w^2}{k_o \phi_b^2 + \phi_w^2} = \frac{\sigma_w^2}{\phi_w^2} \quad (3.57)$$

However, these are now ratios of independent variances, and their estimates a_1/b_1 and a_2/b_2 are each distributed as $F_i \omega_i$, $i = 1, 2$; where the F_i 's follow an F-distribution in which the degrees of freedom are the same for both the numerator and the denominator, and the ω_i 's are as in equations (3.9) and (3.10)..

It should be pointed out that for the case of independent experiments it is not necessary to write the null hypothesis in this form. However, this form of the hypothesis gives us the advantage that the F-ratios to be used in determining the distribution of the test statistic have the same degrees of freedom in both the numerator and denominator. Hence, when we transform to the z-distribution we obtain symmetry about zero. The development in Section 3.6.4 indicates the effect on the calculation of the critical value of not having the numerator and denominator degrees of freedom equal.

3.6.2 Distribution of the test statistic. Once again the discussion of the tests of hypothesis and confidence bounds in Sections 3.2.2.1 and 3.2.2.2 is applicable. However, the test statistic

$$\frac{a_1 b_2}{a_2 b_1} \text{ is now distributed as } \frac{F_{02}}{F_{01}} \left(\frac{\omega_1}{\omega_2} \right), \quad (3.58)$$

where F_{02} follows an F-distribution with (m_2, m_2) degrees of freedom, F_{01} follows an F-distribution with (m_1, m_1) degrees of freedom, and the ω_i 's are as in Section 3.2.2. If we let $W'' = F_{02}/F_{01}$,

$$\frac{a_1 b_2}{a_2 b_1} \text{ is distributed as } W'' , \quad (3.59)$$

under H_0 .

To obtain the distribution of W'' , let

$$\begin{aligned} u'' &= (1/2) \ln(F_{02}/F_{01}) , \\ &= (1/2) \ln(F_{02}) - (1/2) \ln(F_{01}) , \\ &= z_{02} - z_{01} , \end{aligned} \quad (3.60)$$

where z_{01} and z_{02} are distributed as Fisher's z with (m_1, m_1) and (m_2, m_2) degrees of freedom, respectively.

Now, the moments of the z -distribution, when both degrees of freedom are the same, have been presented previously in Section 3.3.2. They are:

$$E(z_{0i}^{2s+1}) = 0 , \quad s = 0, 1, 2, \dots ,$$

and

$$E(z_{0i}^2) = \frac{1}{m_i} + \frac{1}{m_i^2} + \frac{2}{3m_i^3} ,$$

$$E(z_{0i}^4) = \frac{3}{m_i^2} + \frac{8}{m_i^3} + \frac{13}{m_i^4} ,$$

$$E(z_{0i}^6) = \frac{30}{m_i^4} + \frac{111}{m_i^5} + \frac{110}{m_i^6} + \frac{60}{m_i^7} .$$

Hence, the moments of $u'' = z_{02} - z_{01}$ are:

$$E(u''^{2s+1}) = 0 , \quad s = 0, 1, 2, \dots ,$$

and

$$E(u''^2) = \left(\frac{1}{m_1} + \frac{1}{m_2} \right) + \left(\frac{1}{m_1^2} + \frac{1}{m_2^2} \right) + \left(\frac{2}{3m_1^3} + \frac{2}{3m_2^3} \right) ,$$

$$E(u''^4) = \left(\frac{3}{m_1^2} + \frac{3}{m_2^2} \right) + \frac{6}{m_1 m_2} + \left(\frac{8}{m_1^3} + \frac{8}{m_2^3} \right) + \left(\frac{6}{m_1 m_2^2} + \frac{6}{m_2 m_1^2} \right) + \dots ,$$

$$E(u''^6) = \left(\frac{30}{m_1 m_2^2} + \frac{30}{m_2 m_1^2} \right) + \left(\frac{30}{m_1^4} + \frac{30}{m_2^4} \right) + \frac{6}{m_1 m_2} + \dots .$$

Again we see that it is reasonable to drop higher order terms and to use a normal approximation; i.e., $u''/\sigma_{u''}$ is approximately distributed as a standard normal. For m_1 between five and ten use

$$\sigma_{u''}^2 = \left(\frac{1}{m_1} + \frac{1}{m_2}\right) + \left(\frac{1}{m_1^2} + \frac{1}{m_2^2}\right) + \left(\frac{2}{3m_1^3} + \frac{2}{3m_2^3}\right), \quad (3.61)$$

for m_1 between ten and thirty-five use

$$\sigma_{u''}^2 = \left(\frac{1}{m_1} + \frac{1}{m_2}\right) + \left(\frac{1}{m_1^2} + \frac{1}{m_2^2}\right), \quad (3.62)$$

and for m_1 greater than thirty-five use

$$\sigma_{u''}^2 = \left(\frac{1}{m_1} + \frac{1}{m_2}\right). \quad (3.63)$$

As for the tests developed in Sections 3.2 and 3.3, the distribution of W'' will be approximately a lognormal distribution with mean $e^{2\sigma_{u''}^2}$ and variance $e^{4\sigma_{u''}^2}(e^{4\sigma_{u''}^2} - 1)$. Hence, the critical value is

$$W''_{1-\alpha/2} = e^{2\sigma_{u''}^2} \xi_{1-\alpha/2} \quad (3.64)$$

where $\xi_{1-\alpha/2}$ is the normal deviate corresponding to an area under the normal curve of $1-\alpha/2$, and $\sigma_{u''}$ is as above.

Finally, it should be pointed out that the determination of relative sample size presented in Section 3.4 is also applicable to this case.

3.6.3 Comparison of the approximate test to the exact test. In order to compare the approximate test to the Schumann-Bradley test, Table 3.2 has been prepared. This table presents the critical values for both tests for various values of the degrees of freedom. The upper values are from the table presented by Schumann and Bradley (1957) and (1959), while the lower values are the critical values calculated as

Table 3.2. Comparison of critical values for the 5% level of significance for the Schumann-Bradley test (upper value) and the new approximate test (lower value)

$m_1 \backslash m_2$	6	8	10	12	14	16	18	20	30	∞
5	8.87	7.70	7.11	6.68	6.42	6.23	6.09	5.96	5.64	5.05
	8.94	7.77	7.10	6.75	6.42	6.30	6.11	6.05	5.70	5.10
6	7.86	6.77	6.18	5.78	5.54	5.38	5.25	5.13	4.82	4.28
	7.92	6.82	6.23	5.87	5.58	5.42	5.26	5.15	4.86	4.31
8		5.75	5.21	4.88	4.65	4.48	4.35	4.25	3.94	3.44
		5.76	5.21	4.90	4.66	4.48	4.35	4.26	3.97	3.46
9		5.45	4.92	4.59	4.36	4.19	4.07	3.98	3.68	3.18
		5.47	4.95	4.62	4.35	4.18	4.10	3.97	3.71	3.19
10			4.70	4.37	4.14	3.98	3.86	3.76	3.48	2.98
			4.68	4.35	4.14	3.98	3.86	3.74	3.49	2.97
11			4.52	4.19	3.98	3.82	3.70	3.60	3.32	2.82
			4.50	4.18	3.98	3.82	3.71	3.60	3.32	2.83
12				4.05	3.84	3.68	3.56	3.46	3.18	2.69
				4.06	3.82	3.67	3.56	3.46	3.19	2.69
14					3.63	3.47	3.36	3.26	2.98	2.49
					3.63	3.46	3.35	3.25	3.00	2.48
15					3.54	3.39	3.27	3.18	2.90	2.40
					3.52	3.39	3.25	3.19	2.92	2.41
16						3.32	3.20	3.11	2.83	2.34
						3.32	3.20	3.13	2.86	2.34
17						3.26	3.14	3.05	2.77	2.28
						3.25	3.13	3.06	2.77	2.27
18							3.09	3.00	2.72	2.22
							3.10	3.00	2.72	2.22
20								2.91	2.63	2.12
								2.89	2.64	2.12
21								2.87	2.59	2.08
								2.86	2.59	2.08
30									2.36	1.85
									2.36	1.84

above. Since the Schumann-Bradley table presents critical values such that $1 - G(w_0) = .05$, the critical value for the approximate test is

$$W_{.95} = e^{3.29\sigma_u^2}$$

As one might expect, for the smallest values of m_1 , five or six, the value for the approximate test is not as close to the value for the exact test as it is for larger values. However, even for these two values the maximum discrepancy is only $+.09$, while the average discrepancy is $+.046$. For the remainder of the table the maximum discrepancy is $+.04$ and the average discrepancy is $+.0015$. Note that there is no negative discrepancy greater than $.02$. In general then, and particularly for small values of m_1 , we can say that the approximate test is a slightly more conservative test of the null hypothesis of equal sensitivity.

3.6.4 Dissimilar experiments. Occasionally, situations might arise wherein the experiments to be used for comparing the sensitivities of two measuring methods do not have the same degrees of freedom. This might be the case if some experimental units are accidentally destroyed in one experiment and not in the other; or possibly, even though the same treatments are used in each experiment, the experiments are based on different numbers of replications or have different experimental designs.

In this case, neither the degrees of freedom nor the coefficient of the between-grades component of variance will be the same from one experiment to the other. When this is the case, the discussion of the hypothesis H'_0 , presented in Section 3.2.1, will not be applicable. In

that section it was pointed out that the null hypothesis we wished to test was:

$$H_0: \frac{\sigma_b^2}{\sigma_w^2} = \frac{\phi_b^2}{\phi_w^2},$$

and it was shown that, since both k_0 's were the same, this could be written

$$H_0: \frac{k_0 \sigma_b^2 + \sigma_w^2}{\sigma_w^2} = \frac{k_0 \phi_b^2 + \phi_w^2}{\phi_w^2}.$$

If the k_0 's are not the same, it is not possible to rewrite H_0 as H_0' . Hence, for this case, H_0 can not be tested merely by comparing the F-ratios for treatments from the analyses of variance.

This situation has been discussed by Schumann and Bradley (1957) and (1959), and they suggested a procedure for testing H_0 . They suggest that, since H_0 implies $(\sigma_b^2/\sigma_w^2) = (\phi_b^2/\phi_w^2) = \omega_0$, we obtain an estimate of (σ_b^2/σ_w^2) from the analysis of method X_a , and an estimate of (ϕ_b^2/ϕ_w^2) from the analysis of method X_b . They then let q equal the average of these two estimates, and, under H_0 , q is assumed to be an estimate of ω_0 . In the notation of this paper their test statistic would be:

$$\frac{\left(\frac{a_1 m_2}{a_2 m_1}\right)(1 + k_2 q)}{\left(\frac{b_1 m_1^*}{b_2 m_1^*}\right)(1 + k_1 q)} \text{ distributed as } \frac{F_1}{F_2}; \quad (3.65)$$

where (m_1, m_2) and k_1 are the degrees of freedom and coefficient of the between-grades component of variance associated with the analysis of measurement method X_a , (m_1^*, m_2^*) and k_2 are the degrees of freedom and coefficient of the between-grades component of variance associated

with the analysis of measurement method X_b , and F_1 and F_2 each follow an F-distribution with (m_1, m_2) and (m'_1, m'_2) degrees of freedom, respectively.

Following the procedure used previously in this chapter, if we set $W'' = (F_1/F_2)$, and let

$$u'' = (1/2) \ln(W'') \quad , \quad (3.66)$$

we get that $u'' = z_1 - z_2$, where z_1 and z_2 each follow a z-distribution with (m_1, m_2) and (m'_1, m'_2) degrees of freedom, respectively. For this case, z_1 and z_2 are not symmetrically distributed about zero. Hence, using the moments of the general z-distribution as developed by Cornish and Fisher (1937), the mean and variance of u'' will be

$$\mu_{u''} = \frac{1}{2} \left(\frac{1}{m_2} - \frac{1}{m_1} - \frac{1}{m'_2} + \frac{1}{m'_1} \right) + \frac{1}{6} \left(\frac{1}{m_2^2} - \frac{1}{m_1^2} - \frac{1}{m'^2_2} + \frac{1}{m'^2_1} \right),$$

and

$$\begin{aligned} \sigma_{u''}^2 = \frac{1}{2} \left(\frac{1}{m_2} + \frac{1}{m_1} + \frac{1}{m'_2} + \frac{1}{m'_1} \right) + \frac{1}{2} \left(\frac{1}{m_2^2} + \frac{1}{m_1^2} + \frac{1}{m'^2_2} + \frac{1}{m'^2_1} \right) \\ + \frac{1}{6} \left(\frac{1}{m_2^3} + \frac{1}{m_1^3} + \frac{1}{m'^3_2} + \frac{1}{m'^3_1} \right); \end{aligned}$$

Now, since u'' is not symmetrically distributed about zero, we find that $(\mu_{u''} - u'')/\sigma_{u''}$ approximates the standard normal distribution; and the critical value $W''_{1-\alpha/2}$ will now be calculated from

$$W''_{1-\alpha/2} = e^{2\sigma_{u''} \xi_{1-\alpha/2}} \quad , \quad (3.67)$$

where $\sigma_{u''}^2$ and $\mu_{u''}$ are as above, and $\xi_{1-\alpha/2}$ is the normal deviate corresponding to an area under the normal curve of $1-\alpha/2$.

As Schumann and Bradley point out, this is a somewhat crude method of obtaining an approximation to the critical value for this situation, but the suitability of the procedure is not yet subject to check. Where one experiment is conducted and both methods of measurement are applied to the same experimental units, this situation can not arise.

CHAPTER IV

APPLICATION

4.1 Test Procedures

As stated previously, experiments may be carried out specifically for the purpose of comparing the sensitivities of two methods of measurement. When such experiments are conducted, it is assumed that it is possible to select materials which are known to span the range of values for the characteristic of interest; or that it is possible to select treatments such that when they are applied to the material under study they bring about differences in the characteristic of interest. It is not necessary to know the true differences between experimental units, only that differences exist. It is important to note that it is not the effects of the treatments themselves that are of interest, but rather that the differences due to the treatments are of sufficient magnitude that they would be considered to be of practical importance. The treatments then are used only as a means of obtaining a sample of the population we wish to study.

Although, in the tests of sensitivity described in this paper, the statistic W is calculated using sums of squares, W is in reality a ratio of the F -ratios for treatments from the analysis for each measurement method. The most sensitive method is defined as that method which better demonstrates the existence of a "between-treatments" component of variance (Model II of the analysis of variance). Thus, equality of the F -ratios implies equal sensitivities for the two methods of measurement. The null hypothesis representing this equality of sensitivities is

$$H_0^! : \frac{\sigma_w^2 + k_0 \sigma_b^2}{\sigma_w^2} = \frac{\phi_w^2 + k_0 \phi_b^2}{\phi_w^2}, \quad (4.1)$$

where the terms in the ratios are the expected mean squares from the analysis of variance and covariance as illustrated in Table 3.1 and Section 3.2.1. In the theoretical development presented in Chapter III, ω_1 and ω_2 were defined as

$$\omega_1 = \frac{\sigma_w^2 + k_0 \sigma_b^2}{\phi_w^2 + k_0 \phi_b^2}, \quad (4.2)$$

and

$$\omega_2 = \frac{\sigma_w^2}{\phi_w^2}. \quad (4.3)$$

The null hypothesis was then written

$$H_0^! : \omega_1 = \omega_2. \quad (4.4)$$

There are three alternatives against which $H_0^!$ may be tested. They are

$$H_{a1}^! : \omega_1 > \omega_2, \quad (4.5)$$

$$H_{a2}^! : \omega_1 < \omega_2, \quad (4.6)$$

and

$$H_{a3}^! : \omega_1 \neq \omega_2. \quad (4.7)$$

From the theoretical development we can see that the test against the alternative $H_{a2}^!$ is merely the reciprocal of the test against $H_{a1}^!$. Also, since $H_0^!$ implies equality of sensitivities, $H_{a1}^!$ implies that the method associated with X_a is more sensitive than the method associated with X_b , $H_{a2}^!$ implies that the method associated with X_b is the most sensitive, while $H_{a3}^!$ merely implies that the two methods under consideration have different sensitivities.

An α -level test for H_0^I against

- 1) H_{a1}^I will be; "Reject H_0^I if $W > W_{1-\alpha}$, otherwise do not reject H_0^I ."
- 2) H_{a2}^I will be; "Reject H_0^I if $W < 1/W_{1-\alpha}$, otherwise do not reject H_0^I ."
- 3) H_{a3}^I will be; "Reject H_0^I if $W < 1/W_{1-\alpha/2}$ or $W > W_{1-\alpha/2}$, otherwise do not reject H_0^I ."

In the discussion above the test statistic has been referred to as W , the statistic for Case I, however, it should be kept in mind that all comments made in this section are equally applicable to W' , the statistic for Case II, and W'' , the statistic for Case III. The basic difference in the significance tests for each case is whether or not it is necessary to calculate the correlation coefficients, r_1 and r_2 ; i.e., to calculate the critical value for a particular case we must first obtain the appropriate σ_u^2 , $\sigma_{u'}^2$, or $\sigma_{u''}^2$ which, in turn, are dependent on r_1^2 and r_2^2 , only r_1^2 , and neither r_1^2 nor r_2^2 , respectively.

To perform the tests of significance for comparing the sensitivities of two measurement methods the following steps are required.

1. Do the analysis of variance and covariance for the experiment as shown in Table 3.1, Section 3.2.1. (Table 3.1 is reproduced here to facilitate reference.)

Table 3.1. Analysis of variance and covariance

Source	d.f.	Sx_a^2	$Sx_a x_b$	Sx_b^2
Between Grades	m_1	a_1	c_1	b_1
Within Grades	m_2	a_2	c_2	b_2

2. Formulate the null hypothesis, H_0^i , and the appropriate alternative hypothesis, H_{ai}^i .

3. Decide on the significance level, α , to be used

4. Calculate $W = W^I = W^{II} = (a_1 b_2)/(a_2 b_1)$, selecting the appropriate statistic depending on whether the experiment falls into Case I, II, or III.

5. Compute $A_1 = (1-r_1^2)$, using

$$r_1^2 = \frac{c_1^2}{a_1 b_1}, \text{ and } r_2^2 = \frac{c_2^2}{a_2 b_2} \quad (4.8)$$

If the test statistic to be used is W^I , only A_1 need be computed. If the statistic to be used is W^{II} , this entire step may be omitted.

6. Determine the critical value for the particular case being used from

$$W_{1-\alpha} = e^{2\sigma_u \xi_{1-\alpha/2}}$$

$$W_{1-\alpha}^I = e^{2\sigma_{u^I} \xi_{1-\alpha/2}} \quad (4.9)$$

or

$$W_{1-\alpha}^{II} = e^{2\sigma_{u^{II}} \xi_{1-\alpha/2}}$$

where $\xi_{1-\alpha}$ is the normal deviate corresponding to an area under the normal curve of $1-\alpha$, and σ_u^2 is determined according to the case under consideration as follows:

a) For Case I

$$\sigma_u^2 = \left(\frac{A_1}{m_1} + \frac{A_2}{m_2} \right) + \left(\frac{3A_1 - A_1^2}{m_1^2} + \frac{3A_2 - A_2^2}{m_2^2} \right) + \left(\frac{9A_1 - 8A_1^2 + 8/3 A_1^3}{m_1^3} + \frac{9A_2 - 8A_2^2 + 8/3 A_2^3}{m_2^3} \right) \quad (4.10)$$

b) For Case II

$$\sigma_{u'}^2 = \left(\frac{A_1}{m_1} + \frac{1}{m_2} \right) + \left(\frac{3A_1 - A_1^2}{m_1^2} + \frac{1}{m_2} \right) + \left(\frac{9A_1 - 8A_1^2 + 8/3 A_1^3}{m_1^3} + \frac{1}{m_2} \right) . \quad (4.11)$$

c) For Case III

$$\sigma_{u''}^2 = \left(\frac{1}{m_1} + \frac{1}{m_2} \right) + \left(\frac{1}{m_1^2} + \frac{1}{m_2} \right) + \left(\frac{2}{3m_1^3} + \frac{2}{3m_2} \right) . \quad (4.12)$$

For values of m_1 from five to ten calculate the variances as shown above; for m_1 between ten and thirty-five the third bracketed term on the right may be dropped before calculating the variances; and, for m_1 equal to or greater than thirty-five the second and third bracketed terms may be dropped.

7) Compare the test statistic with the appropriate critical value and reject H_0 in favor of H_{a_i} in accordance with the rules set forth previously.

Confidence bounds may be obtained for ω_1/ω_2 by the relation

$$\text{Prob.} \left\{ \frac{a_1 b_2}{a_2 b_1} / W_{1-\alpha/2} \leq \omega_1/\omega_2 \leq \frac{a_1 b_2}{a_2 b_1} \cdot W_{1-\alpha/2} \right\} = 1 - \alpha . \quad (4.13)$$

This relation will hold for all three cases by using the appropriate critical value.

Finally, the relative sample size for either of the three cases may be calculated from the following relations.

a) For the relative sample size of method X_b given the sample size to be used with method X_a :

$$h_{ba} = \frac{\left(\frac{a_1}{m_1} - \frac{a_2}{m_2} \right)}{\left(\frac{b_1}{m_1} - \frac{b_2}{m_2} \right)} \left(\frac{b_2}{a_2} \right) , \quad (4.14)$$

b) For the relative sample size of method X_a given the sample size to be used with method X_b :

$$h_{ab} = \frac{\left(\frac{b_1}{m_1} - \frac{b_2}{m_2}\right)}{\left(\frac{a_1}{m_1} - \frac{a_2}{m_2}\right)} \left(\frac{a_2}{b_2}\right) \quad (4.15)$$

4.2 Example of a Case I Experiment

The following example is taken from a study by Fromm (1959). The objective of the study was to compare two methods for determination of egg shell permeability. The first method utilized a dye penetration technique and a measure of optical density, while the second was a measure of the egg weight loss. It was known that the permeability of the egg shell increases with age so that nine ages were selected as the treatments, and eight pairs of eggs were observed for each age. The weight observations on each pair were averaged, while the optical density reading was made on a pooled sample of the shells from each pair. Hence, there was only one optical density reading per pair.

The test for comparison of the sensitivities of the two methods now follows:

1. The analysis of variance and covariance is given in Table 4.1.

Table 4.1. Analysis of variance and covariance for egg data

Source	d.f.	Sx_a^2	$Sx_a x_b$	Sx_b^2
Between Ages	8	.7896	7.2760	81.3923
Within Ages	63	.8194	.3189	6.0666

X_a =Optical density reading

X_b =Weight loss determination

2. The hypothesis to be tested is $H_0^1 : \omega_1 = \omega_2$, against $H_{a3}^1 : \omega_1 \neq \omega_2$.

3. It is specified that $\alpha = .05$.

$$4. W = \frac{.7896}{.8194} \times \frac{6.0666}{81.3923} = .0718 .$$

$$5. r_1^2 = \frac{(7.2760)^2}{(.7896)(81.3923)} = .8237 \quad , \quad A_1 = .1763 .$$

$$r_2^2 = \frac{(.3189)^2}{(.8194)(6.0666)} = .0205 \quad , \quad A_2 = .9795 .$$

$$6. \sigma_u^2 = \frac{.1763}{8} + \frac{.9795}{63} + \frac{3(.1763) - (.1763)^2}{64} + \frac{3(.9795) - (.9795)^2}{3969} ,$$

$$= .0468337 .$$

$$\sigma_u = .2164 .$$

The third term of σ_u^2 was not used here, even though $m_1 < 10$, since it was obviously less than .0001

$$W_{.975} = e^{3.92(.2164)} = 2.3396 \quad ,$$

$$W_{.025} = e^{-3.92(.2164)} = .4274 \quad .$$

7. Since $W < W_{\alpha/2}$, H_0^1 is rejected in favor of H_{a3}^1 ; i.e., The results of this experiment indicate that there is a significant difference in the sensitivities of the two methods. Obviously, the weight loss procedure is the most sensitive procedure.

A 95% confidence interval for the sensitivity ratio is given by

$$W \times W_{.025} \leq \omega_1/\omega_2 \leq W \times W_{.975} ;$$

that is,

$$(.0718)(.4274) \leq \omega_1/\omega_2 \leq (.0718)(2.3396) \quad ,$$

$$.0307 \leq \omega_1/\omega_2 \leq .1680 \quad .$$

Or, since the method corresponding to X_p was the most sensitive method, we might prefer the confidence interval for ω_2/ω_1 . This is

$$W_{.025}/W \leq \omega_2/\omega_1 \leq W_{.975}/W \quad ,$$

which gives

$$5.953 \leq \omega_2/\omega_1 \leq 32.585 \quad .$$

4.3 Example of a Case II Experiment

The following example is taken from a study presented by Hart (1958). The objective of the study was to develop a more rapid test method for determining the durability of Type II (water-resistant) hardwood plywood. Since the standard test requires up to fifteen days for acceptance or rejection of a shipment, while the proposed accelerated procedure requires only twenty-four hours for a determination of durability, it was of interest to determine which test does the better job of measuring durability. For the comparison of sensitivities the standard test was modified. Normally the test is run for a maximum of fifteen cycles of wetting and drying, and the cycle at which two inches of continuous delamination along any glue line of the test specimen occurs is recorded. The specimens are required to average ten cycles or better to prove acceptable. For this study the test was continued beyond fifteen cycles, and the cycle at which the required delamination finally occurred was recorded. The new test, which introduced a new type of test specimen as well as an accelerated wetting-drying cycle, was run for twenty-four cycles and the measure of durability used was the percent of the exposed edges on which delamination had occurred.

For the portion of the study presented here, three glues were selected, two urea resin glues and one protein glue. These glues will be referred to as UF65-100, UF65-150, and soybean. It was of interest to compare the sensitivities both within the various glues, and over all three glues. The treatments used to effect differences in durability were combinations of species of wood, veneer thickness, and construction (ply). In all, sixteen panels were prepared for each glue mixture, and two specimens from each panel were tested by each procedure. For the comparison of sensitivities, those panels whose specimens failed on the first cycle with the standard procedure and also registered 100% delamination on the accelerated procedure were dropped. Hence, for the final analysis there were less than sixteen panels for two of the glues.

For the analysis of the data, a logarithmic transformation has been applied to the results of the standard procedure, X_a ; while an arcsine transformation has been applied to the results of the accelerated procedure, X_p .

The test for the comparison of the sensitivities of the standard test and the accelerated test is given below.

1. Since it is of interest to compare the sensitivities for the three glues separately, Table 4.2 presents the analysis for each of the glues.

Table 4.2. Analysis of variance and covariance for the three glue mixtures.

Source	d.f.	Sx_a^2	$Sx_a x_b$	Sx_b^2
UF65-100				
Between Panels	13	4.0879	-7.4312	16.5562
Within Panels	14	.3872		3.6788
UF65-150				
Between Panels	15	4.5834	-6.1242	17.7663
Within Panels	16	.6377		2.9139
Soybean				
Between Panels	10	1.6985	-2.4373	5.3969
Within Panels	11	.8705		4.1896

2. The null hypothesis is $H_0^1 : \omega_1 = \omega_2$, and the alternative hypothesis is $H_{a3}^1 : \omega_1 \neq \omega_2$. The two-sided alternative is used here since the new procedure is twelve to fifteen times faster than the standard procedure; and, hence, might be used even if it is less sensitive, depending on how much less sensitive it is.

3. The significance level to be used is $\alpha = .05$.

For UF65-100:

$$4. S^1 = \frac{4.0879}{16.5562} \times \frac{3.6788}{.3872} = 2.358 \quad .$$

$$5. r_1^2 = \frac{(-7.4312)^2}{(4.0879)(16.5562)} = .8159 \quad , \quad A_1 = .1841 \quad .$$

$$6. \sigma_{u^1}^2 = \frac{.1841}{13} + \frac{1}{14} + \frac{3(.1841) - (.1841)^2}{169} + \frac{1}{196} = .09378 \quad .$$

$$\sigma_{u^1} = .306 \quad .$$

Thus,

$$W_{.975}^t = e^{3.92(.306)} = 3.3201 ,$$

$$W_{.025}^t = e^{-3.92(.306)} = .3012 .$$

For UF65-150:

$$4. W^t = \frac{4.5834}{17.7663} \times \frac{2.9139}{.6377} = 1.1789 .$$

$$5. r_1^2 = \frac{(-6.1242)^2}{(4.5834)(17.7663)} = .4606 , A_1 = .5394 .$$

$$6. \sigma_{u^t}^2 = \frac{.5394}{15} + \frac{1}{16} + \frac{3(.5394) - (.5394)^2}{225} + \frac{1}{256} = .10826 .$$

$$\sigma_{u^t} = .329 .$$

Thus,

$$W_{.975}^t = e^{3.92(.329)} = 3.6328 ;$$

$$W_{.025}^t = e^{-3.92(.329)} = .2753 .$$

For Soybean:

$$4. W^t = \frac{1.6985}{5.3969} \times \frac{4.1896}{.8705} = 1.5146 .$$

$$5. r_1^2 = \frac{(-2.4373)^2}{(1.6985)(5.3969)} = .6480 , A_1 = .3520 .$$

$$6. \sigma_{u^t}^2 = \frac{.352}{10} + \frac{1}{11} + \frac{3(.352) - (.352)^2}{100} + \frac{1}{121} = .14368 ,$$

$$\sigma_{u^t} = .379 .$$

Thus,

$$W_{.975}^t = e^{3.92(.379)} = 4.4371 ,$$

$$W_{.025}^t = e^{-3.92(.379)} = .2254 .$$

7. For all three glues we do not reject H_0^1 . Hence, the results of this experiment indicate that there is no significant difference between the sensitivities of the new accelerated test and the standard test.

Since the sensitivities of the two methods are essentially the same for all three glues, we will now pool the data to obtain an estimate of the sensitivity ratio for all three glues. The analysis of the pooled data is as follows:

1. Table 4.3. Analysis of variance and covariance for the pooled plywood data

Source	d.f.	Sx_a^2	$Sx_a x_b$	Sx_b^2
Between Panels	40	11.9338	-17.3624	41.0982
Within Panels	41	1.8954		10.7723

$$2. W^* = \frac{11.9338}{41.0982} \times \frac{10.7723}{1.8954} = 1.6511 \quad ,$$

$$3. r_1^2 = \frac{(-17.3624)^2}{(11.9338)(41.0982)} = .6416 \quad , \quad A_1 = .3854 \quad .$$

$$4. \sigma_u^2 = \frac{.3854}{40} + \frac{1}{41} = .034025 \quad ,$$

$$\sigma_u = .184 \quad .$$

Thus,

$$W_{.975}^* = e^{3.92(.184)} = 2.0544$$

$$W_{.025}^* = e^{-3.92(.184)} = .4868 \quad .$$

Then confidence limits for the overall sensitivity ratio are

$$.8038 \leq \omega_1 / \omega_2 \leq 3.3920 \quad .$$

Even if the results of an experiment to compare the sensitivities of two methods of measurement indicate that there is no difference in sensitivity between the two methods, we may still wish to calculate relative sample size. For this particular example one would not need to determine relative sample size. However, in order to demonstrate the computations involved, the determination of the relative sample size of method X_b , given that two samples are to be observed if method X_a is used, is as follows:

$$h_{ba} = \frac{(.2983 - .0462)}{(1.0274 - .2627)} \times \frac{41.0982}{11.9338} = 1.135 .$$

Hence, the sample size to be used with the new method to insure that the same degree of sensitivity is attained will be $(1.135) \times (2) = 2.27$ or 2. This is as expected for this example.

CHAPTER V

SUMMARY

5.1 Summary

In this dissertation statistical tests of significance are developed for comparing the sensitivities of two methods of measuring some characteristic or response of interest, where both measuring methods are applied to the same experiment. Two experimental situations were of primary interest, namely:

- (i) Case I: Both measurement methods applied to the same experimental units,
- (ii) Case II: Each measurement method applied to separate sub-samples from the same experimental units.

A technique is presented for determining the relative sensitivity of the two measurement methods. The relative sensitivity is then used to determine the number of additional samples required with the least sensitive method to provide the same sensitivity as would be obtained with the most sensitive method.

The power of the tests developed for Cases I and II is discussed, and power curves are presented for several values of the parameters.

As an outgrowth of the study for Case I and Case II, a test was developed for a third experimental situation; viz.,

- (iii) Case III: Each measurement method applied to independent experiments.

Two situations are dealt with; the first is the case where both experiments have the same degrees of freedom, and the second is the case where

the degrees of freedom are different in each experiment. A table is presented comparing critical values for the 5% significance level of the test developed in this study, and the corresponding critical values of a test developed previously by Schumann and Bradley.

Illustrations are given in Chapter IV to present the computational procedures for use of the tests for Case I and Case II. Confidence bounds are calculated for the sensitivity ratios, and the use of the relative sample size for determining required sample size is demonstrated.

5.2 Suggestions for Further Research

The development in this dissertation has been restricted to Model II (the random effects model) of the analysis of variance. There is a need for the development of a statistical test for the comparison of sensitivities of two measurement methods applied to the same experiment when Model I (the fixed effects model) of the analysis of variance is the applicable model.

An extension of the procedures presented here for comparing more than two measurement methods applied to the same experiment is also needed. For this case some form of ranking procedure will be necessary.

Finally, the criterion used here selects as the most sensitive method that method which exhibits the optimum combination of maximum discrimination of differences in the characteristic of interest and maximum repeatability. In some situations it seems reasonable to assume that what may be required is some weighting of these two factors, either in favor of more precision or in favor of greater ability to discriminate. Hence, a critical study of the criterion for the most sensitive measurement method would be a worthwhile investigation.

LIST OF REFERENCES

- Bradley, R. A. and Schumann, D. E. W. 1957. The comparison of sensitivities of similar experiments: Applications. *Biometrics*, 13: 496-510.
- Bross, I. D. J. 1959. Note on an application of the Schumann-Bradley Table. *Ann. Math. Stat.* 30:581-583.
- Carlin, A. F., Kempthorne, O., and Gordon, J. 1956. Some aspects of numerical scoring in subjective evaluation of foods. *Food Research*, 21:273-281.
- Cochran, W. G. 1943. The comparison of different scales of measurement for experimental results. *Ann. Math. Stat.* 14: 205-216.
- Cornish, E. A. and Fisher, R. A. 1937. Moments and cumulants in the specification of distributions. *Extrait de la Revue de l'Institut International de Statistique*, 4:1-14.
- Cramér, H. 1957. Mathematical Methods of Statistics. Princeton University Press. Princeton, N. J.
- Dillon, J. H. 1936. A simplified extrusion plastometer for unvulcanized rubber. *Physics*, 7:73-76.
- Eisenhart, C. 1947. The assumptions underlying the analysis of variance. *Biometrics*, 3:1-21.
- Fromm, D. 1959. A rapid method for the determination of egg shell permeability. *Poultry Sci.* 38:171-173.
- Grubbs, F. E. 1948. On estimating precision of measuring instruments and product variability. *Jour. Amer. Stat. Assoc.* 43:243-264.
- Hart, C. A. 1958. Development of a highly accelerated test for water-resistant hardwood plywood. *For. Prod. Jour.* 8:113-117.
- Lashof, T. W., Mandel, John, and Worthington, V. 1956. Use of the sensitivity criterion for the comparison of the Bekk and Sheffield smoothness testers. *TAPPI*, 39:532-543.
- Mandel, J. and Stiehler, R. D. 1954. Sensitivity — a criterion for the comparison of methods of test. *Jour. Res. Nat'l Bur. Stand.* 53:155-159.
- Mandel, John. 1959. The measuring process. *Technometrics*, 1:251-267.

- Mandel, J. and Lashof, T. W. 1959. The interlaboratory evaluation of testing methods. ASTM Bulletin, No. 239:53-61.
- Patnaik, P. B. 1949. The non-central X^2 - and F-distributions and their applications. Biometrika, 36:202-232.
- Roth, F. L. and Stiehler, R. D. 1948. Strain test for evaluation of rubber compounds. Jour. Res. Nat'l Bur. Stand. 41:87-93.
- Schumann, D. E. W. and Bradley, R. A. 1957. The comparison of the sensitivities of similar experiments: Theory. Ann. Math. Stat. 28:902-920.
- Schumann, D. E. W. and Bradley, R. A. 1959. The comparison of the sensitivities of similar experiments: Model II of the analysis of variance. Biometrics, 15:405-416.
- Smith, H. F. 1950. Estimating precision of measuring instruments. Jour. Amer. Stat. Assoc. 45:447-451.
- Teichman, R., Beall, G., Eaton, H. D., Rousseau, J. E., Jr., Dolge, K.L., Moore, L. A., and Frey, P. R. 1957. Sensitivity of blood plasma and liver vitamin A concentrations and vitamin A depletion time to carotene and/or vitamin A intake. Jour. Dairy Sci. 40:1284-1293.