INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in “sectioning” the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from “photographs” if essential to the understanding of the dissertation. Silver prints of “photographs” may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

University Microfilms International
300 North Zeib Road
Ann Arbor, Michigan 48106 USA
St. John's Road, Tyler's Green
High Wycombe, Bucks, England HP10 8HR
BISHOP, Thomas Alan, 1949-
HETEROSCEDASTIC ANOVA, MANOVA, AND
MULTIPLE-COMPARISONS.

The Ohio State University, Ph.D., 1976
Statistics

Xerox University Microfilms, Ann Arbor, Michigan 48106
HETEROSCEDASTIC ANOVA, MANOVA, AND
MULTIPLE-COMPARISONS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By
Thomas Alan Bishop, B.S., M.A.

The Ohio State University
1976

Reading Committee:
Dr. Edward J. Dudewicz
Dr. D. Ransom Whitney
Dr. Jagdish S. Rustagi
Dr. Douglas A. Wolfe

Approved by

Edward J. Dudewicz
Adviser
Department of Statistics
To My Wife Vicki
VITA

June 14, 1949 . . . . . . Born - Columbus, Ohio

1971 . . . . . . . . . . . . . . B.S., Ohio University, Athens, Ohio

1973 . . . . . . . . . . . . . M.A., Indiana University, Bloomington, Indiana

1973-1976 . . . . . . Teaching Associate, Department of Statistics, The Ohio State University, Columbus, Ohio

FIELDS OF STUDY

Major Field: Statistics

Studies in Ranking and Selection. Professor Edward J. Dudewicz

Studies in Analysis of Variance. Professor Peter O. Anderson

Studies in Multivariate Analysis. Professor Jagdish S. Rustagi

Studies in Statistical Inference. Professor Douglas A. Wolfe
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGMENTS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>VITA</td>
<td>iii</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter</th>
<th>GOALS, HISTORY AND LITERATURE</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1.1 Goals</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1.2 Notation and History</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1.3 Review of the Literature</td>
<td>18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II</th>
<th>ANOVA</th>
<th>89</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.1 ANOVA in an r-Way Layout</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>2.2 Multiple-Comparison Procedures</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>2.2A All Pairwise Differences</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>2.2B Contrasts</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td>2.2C All Linear Combinations: Augmented Range</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>2.2D All Linear Combinations: Scheffé Type</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>2.2E Contrasts: Scheffé Type</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>2.3 ANOVA in a Two-Way Layout</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>2.4 ANOVA in a One-Way Layout $P_{B_1}$</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>2.5 Procedure $P_{B_2}$</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>2.6 Procedure $P_{B_3}$</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>2.7 Comparison of Procedures $P_{B_1}$, $P_{B_2}$, $P_{B_3}$</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>2.8 Comparison of Procedures Based on $X$'s with Procedures Based on $X$'s</td>
<td>161</td>
</tr>
<tr>
<td></td>
<td>2.9 A Note on the Transformation of One-Stage Procedures into Two-Stage Procedures</td>
<td>170</td>
</tr>
</tbody>
</table>
### III MONTE CARLO SAMPLING STUDIES

3.1 Design of the Monte Carlo Study .......................... 175
3.2 The Distribution of $\bar{F}$ (Tables of Level and Power of $B_1$) ............................................ 177
3.3 Study of the Limiting Approximations to the Distribution of $F'$ .................................... 194
3.4 Non-robustness of the $F$-Test ..................... 199
3.5 Relative Efficiency of $F$ and $F'$ ............... 212

### IV HETEROSCEDASTIC DECISION THEORY AND THE HETEROSCEDASTIC METHOD ................. 215

4.1 The Statistical Decision Problem .................... 215
4.2 The Two-Stage Sampling Procedure .......... 217
4.3 Examples of Some Statistical Decision Problems .................................................. 224
4.4 Nonexistence of Single-Sample Procedures 231
4.5 Limiting Distribution of $X$ ......................... 235

### V MANOVA ................................................................. 237

5.1 Introduction .................................................. 237
5.2 The One-Way MANOVA Layout ................. 238
5.3 The Limiting Distribution of $\bar{F}$ ............... 243
5.4 Multivariate Multiple-Comparisons—One Population .................................................. 248
5.5 Multiple-Comparison Procedures with MANOVA .................................................. 251

APPENDIX A. Computer Programs and Examples of Output ........................................... 254

BIBLIOGRAPHY .................................................. 269
1.1 Goals

The theory of analysis of variance under the (classical) assumptions of normality, independence, and homoscedasticity of errors is quite complete. Studies of the robustness of the F-statistic, for example in the one-way layout, have shown that the violation of normality has little effect on inferences about the means. However, the violation of independence or of equality of variances can have a serious effect on inferences about the means, especially if the cell sizes are unequal (see, for example, Scheffé (1959), Brown and Forsythe (1974a), Box (1954a), Chakravarti (1965), or Box (1954b)).

In practice the assumption of equality of error variances seems to be the assumption most often unjustified, and at the same time there is a serious gap in the theory under conditions of heteroscedasticity for testing hypotheses and multiple-comparisons. Historically transformations of the data, for example logarithm or arcsine transformations, have been used to attempt to achieve equality of variances. While these techniques are still in widespread use today, they are only approximate in terms of equal variances and normality assumptions, and often introduce as many complications as they attempt to eliminate, as we will note in detail below. Many of the usual transformations are special cases of the following general transformation. Suppose that \( Y \) is a random
variable with mean \( \mu \) and that the standard deviation of \( Y \) is
\[ \sigma_Y = \varphi(\mu) \]
where \( \varphi \) is a known function. If \( Z = f(Y) \) then the
standard deviation of \( Z \), \( \sigma_Z \), is approximated by
\[ \sigma_Z = \sigma_Y f'(\mu) \]
and we may seek to find \( f \) such that
\[ \sigma_Y f'(\mu) = k \]
where \( k \) is a fixed constant. The solution we seek is then given by
\[ f(Y) = \sigma_Y \int \frac{dy}{\varphi(y)} \]  
(1.1.1)

Problems with this approach are numerous, some of the most obvious being that:

1) We must assume that \( \sigma_Y = \varphi(\mu) \) with \( \varphi \) known;

2) We must assume that \( E(Z) = f(E(Y)) = f(\mu) \) (which in general
is not true); and

3) Multiple-comparisons usually become nonsensical. For example,
in an agricultural experiment designed to study the difference
\( \mu_1 - \mu_2 \) where \( \mu_1 \) and \( \mu_2 \) are respective mean yields of
crops one and two, \( \mu_1 - \mu_2 \) has meaning. Once a transformation
is made, say arcsine, the experimenter is studying arcsine(\( \mu_1 \))-  
arc sine(\( \mu_2 \)), which has no physical meaning to the experimenter.

Thus transformations in general do not provide a satisfactory
solution to the problem of unequal error variances (unless the quantity
observed is \( g(Y) \) where \( Y \) is normally distributed, in which case a
\( g^{-1} \) transformation may be of interest if the observed quantity is not
of intrinsic interest itself).

In recent years several procedures have been proposed to deal
with inequality of variances in the one-way layout and, more generally,
in the general linear model. However, most of these procedures have only been approximate: the exact distribution of the test statistic is not obtained, but rather an approximation to the true distribution is used. More recently, however, problems of heteroscedasticity in ranking and selection have been solved, as have certain multiple-comparison problems.

Even when the error variances are equal the F-statistic for the one-way layout and for the general linear model has one major drawback, its power function depends upon the unknown variance. (For example, in the one-way layout the power depends upon the parameter $\Delta = \sum_{i=1}^{k} n_i (\mu_i - \bar{\mu})^2/\sigma^2$, where $\mu_i$ is the mean of the $i$th population, $n_i$ is the sample-size from the $i$th population and $\bar{\mu} = (1/k) \sum_{j=1}^{k} \mu_j$ ($i=1,2,\ldots,k$).) Because of this fact an experimenter cannot fully design his experiment; he cannot specify sample-sizes which will guarantee specified power at a specified alternative.

Our goals include: development of tests of hypotheses and multiple-comparison procedures for the usual $r$-way analysis of variance (ANOVA) set-up which are completely free of the unknown, unequal variances; development of tests of hypotheses and multiple-comparison procedures for the one-way multivariate analysis of variance (MANOVA) set-up which are completely free of the unknown, unequal covariance matrices; and finally development of decision rules for the general statistical decision problem with risk functions completely independent of the unknown, unequal covariance matrices.
1.2 Notation and History

In this section we set up the standard notation to be used throughout and provide a brief history of testing, multiple-comparisons, and related problems under conditions of heteroscedasticity. (A detailed review of the literature is given in Section 1.3.)

The notations \( t_n \), \( \chi^2_n \), and \( F_{m,n} \) are used as generic symbols for random variables which have a Student's-t distribution with \( n \) degrees of freedom, a chi-square distribution with \( n \) degrees of freedom, and an \( F \)-distribution with \( m \) and \( n \) degrees of freedom, respectively. \( t^\alpha_n \), \( \chi^2_{n;\alpha} \), and \( F^\alpha_{m,n} \) denote the upper \( \alpha \)th percent points of the respective distributions, and \( t_n(\Delta) \), \( \chi^2_n(\Delta) \), \( F_{m,n}(\Delta) \) denote the respective noncentral forms of the distributions with noncentrality parameter \( \Delta \).

The symbol \( N_p(\mu,\Sigma) \) is used to denote the \( p \)-dimensional multivariate normal distribution with mean vector \( \mu = (\mu_1,\ldots,\mu_p)' \) and covariance matrix \( \Sigma \). For the univariate case \( N(0,1) \) we denote the cumulative distribution function by \( \Phi(\cdot) \) and the density function by \( \varphi(\cdot) \). If \( X \) is a random variable which has a certain distribution, for example a Student's-t distribution with \( n \) degrees of freedom, then we denote this by \( X \sim t_n \).

We use the following form for the density of a noncentral chi-square distribution with \( n \) degrees of freedom and noncentrality parameter \( \Delta \) (given, e.g., in Johnson and Kotz (1970)):

\[
  f(x) = e^{-\Delta/2} \sum_{j=0}^{\infty} \frac{(\Delta/2)^j}{\Gamma(n/2+j)} \frac{x^{n/2+j-1}e^{-x/2}}{j!}, \quad 0 < x < \infty. \ 
\]
It should be noted that several equivalent forms of the noncentral chi-square density appear in the literature and one must be careful to note which form he is working with. For example $\Delta = 2\tau$ if $\tau$ is the noncentrality parameter for the form of the noncentral chi-square distribution in Dudewicz (1976).

We denote the $k \times k$ matrix $A$ by $[a_{ij}]$, $A^{-1}$ by $[a_{ij}]^{-1}$, and the diagonal matrix with 0's off the main diagonal and $i$th diagonal element $a_i$ by $\text{diag}(a_1, a_2, \ldots, a_k)$.

The concept of an idempotent matrix will be used extensively as will the notion of the direct product of two matrices.

**Definition (1.2.2):** A matrix $A$ is idempotent if $A \cdot A = A$ (i.e. if $A^2 = A$).

**Definition (1.2.3):** If $A$ is a $q \times q$ matrix $[a_{ij}]$ and $B$ is a $p \times p$ matrix $[b_{lm}]$ then the direct product $A \otimes B$ of $A$ and $B$ is the $pq \times pq$ matrix

$$
\begin{bmatrix}
    a_{11}B & a_{12}B & \cdots & a_{1q}B \\
    \vdots & \vdots & & \vdots \\
    a_{q1}B & a_{q2}B & \cdots & a_{qq}B
\end{bmatrix}
$$

The symbol $\delta_{ij}$ will denote Kronecker's delta and is defined by

$$
\delta_{ij} = \begin{cases} 
0 & i \neq j \\
1 & i = j.
\end{cases}
$$

Dantzig (1940), using the concept of similar regions, proved that there does not exist a single-sample procedure for testing Student's
Hypothesis (i.e. the hypothesis that the mean of a normal population has a specified value when nothing is assumed about the variance) which has a power function independent of the unknown variance. Using similar regions and generalized polar coordinates, he showed that any single-sample test with power independent of the variance had to have a constant power function equal to the level of the test. Hence, only the trivial test has a power function independent of the variance. It should be noted that in practice the more important question is whether there exists a single-stage test procedure with power $\geq \beta$ at $\mu = \mu_1$ for all $\sigma^2$. The nonexistence of such a procedure follows directly from Stein's proof (see below).

Stein (1945) extended this theorem by using a proof involving the most powerful (Neyman-Pearson) test, and showed that no single-sample test of a linear hypothesis for the general linear model (as defined in Scheffé (1959), p.4) exists which is independent of the variance.

These results have important implications for the design of experiments. One of the most important aspects of experimental design is control of the probability that the experiment will reject the null hypothesis given that it is false. Thus, one wants to design his experiment so that the probability of rejecting the null hypothesis is high over a certain (important) subset of alternatives. The problem with single-sample procedures is that their power functions depend upon the unknown variance. Therefore, it is difficult to rationally plan and design experiments based on single-sample procedures since the power cannot be controlled. In fact the results of Dantzig and Stein show
that such a design is impossible and, therefore, by implication, show the
necessity of considering multistage sampling schemes.

Stein (1945) made a historic breakthrough in this area (for con-
trol of a single unknown variance) by developing a two-stage test (for
the general linear hypothesis) whose power is independent of the unknown
variance. Let \( \pi_1, \pi_2, \ldots, \pi_k \) be \( k \) populations such that an observation
from the \( i \)th population \( \pi_i \) is normally distributed with mean \( \mu_i \) and
variance \( \sigma_i^2 \) (denote this association by \( \pi_i \sim N(\mu_i, \sigma_i^2) \)). Stein assumed
\( \sigma_i^2 = \sigma^2 \) and \( \mu_i = \sum_{j=1}^{k'} \beta_j x_{ij} \) where the \( x_{ij} \) are fixed known constants,
\( \sigma^2 \) and \( \{\beta_j\} \) being unknown (\( i = 1, 2, \ldots, k; j = 1, 2, \ldots, k'; k' \leq k \)). He
considered the general linear hypothesis

\[
H_0 : \sum_{j=1}^{k'} c_{\ell j} \xi_j = c_\ell, \quad \ell = 1, 2, \ldots, r \leq k' \tag{1.2.4}
\]

and reduced this problem to canonical form so that

\[
\mu_i = \begin{cases} 
0, & \text{if } i = k'+1, \ldots, k \\
\xi_i', & \text{if } i = 1, 2, \ldots, k'. 
\end{cases} \tag{1.2.5}
\]

The general linear hypothesis then becomes

\[
H_0 : \xi_i = 0, \quad i = 1, 2, \ldots, k'' \leq k' \tag{1.2.6}
\]

(the remaining \( \xi_i \) (\( i = k''+1, \ldots, k' \)) and \( \sigma^2 \) being nuisance parameters).

He then developed a statistic, based on the following two-stage sampling
scheme, for testing (1.2.6) such that the power of the test is indepen-
dent of the unknown variance \( \sigma^2 \). We denote Stein's procedure by \( \Psi_s \).
Procedure \( \mathcal{P} \): Take an initial sample \((Y_{i1}, \ldots, Y_{in_0})\) of size \(n_0\) from population \(\pi_i\) \((i = 1, 2, \ldots, k)\) and estimate \(\sigma^2\) by

\[
s^2 = \frac{1}{(n_0k - k')} \left[ \sum_{j=1}^{n_0} \sum_{i=1}^{k} Y_{ij}^2 - \frac{1}{n_0} \sum_{i=1}^{k} \left( \sum_{j=1}^{n_0} Y_{ij} \right)^2 \right]. \tag{1.2.7}
\]

Let \(z > 0\) be a (prespecified) constant and define

\[
N = \max\{n_0 + 1, \left\lceil \frac{s^2}{z} \right\rceil + 1\}, \tag{1.2.8}
\]

where \([x]\) denotes the greatest integer smaller than \(x\). Note that there is a misprint in the definition of \([x]\) in Stein's 1945 paper and also that some authors, e.g., Dudewicz and Dalal (1975), define \([x]\) to be the smallest integer greater than or equal to \(x\). Choose a set of real numbers \(a_1, a_2, \ldots, a_N\) such that

\[
a_1 = \cdots = a_{n_0}, \tag{1.2.9a}
\]

\[
\sum_{j=1}^{N} a_j = 1, \tag{1.2.9b}
\]

\[
s^2 \sum_{j=1}^{N} a_j^2 = z. \tag{1.2.9c}
\]

Then the statistic

\[
F' = \frac{\sum_{i=1}^{k''} \left( \sum_{j=1}^{N} a_j Y_{ij} \right)^2}{z(n_0k - k') n_0k - k'} \tag{1.2.10}
\]

has a noncentral distribution given by (1.2.13) with \((n_0k-k')\) and \(k''\) degrees of freedom and noncentrality parameter.
where \( F'_{k'', n_0 k-k'; \alpha} \) is the upper \( \alpha \)th percent point of the distribution given by (1.2.13) with \( \Delta = 0 \) (i.e. the central distribution for \( F' \)).

Properties of \( \phi_s \): The power function of the test based on \( F' \) is given by

\[
\beta(s_1, \ldots, s_k) = 1 - \phi_{k'', n_0 k-k'}(F'_{k'', n_0 k-k'; \alpha} \sum_{i=1}^{k''} \xi_i^2 / (z(n_0 k-k')))
\]

where

\[
\phi_{m,n}(T, \Sigma c_i^2) = \frac{\Gamma(m+n/2)}{\sqrt{\pi} \Gamma(m/2) \Gamma(n/2)} \frac{T}{\sqrt{F'}} \int_0^{\sqrt{F'}} (F' - \rho^2) (m-3)/2 \left[ 1 + F' + 2\rho \sqrt{\Sigma c_i^2 + \Sigma c_i^2} \right]^{-(m+n)/2} d\rho dF'.
\]

(1.2.13) is independent of \( \sigma^2 \) (and hence so is the power of the test).

When \( \Sigma c_i^2 = 0 \) the distribution of \( F' \) is that of the ratio

\[
\chi^2_{k''/n_0 k-k'} \quad \text{where} \quad \chi^2_{k''} \quad \text{and} \quad \chi^2_{n_0 k-k'} \quad \text{are independent}.
\]

In the special case of testing Student's Hypothesis, the test statistic \( F' \) reduces to \( t' \).
where \((X_1, \ldots, X_N)\) is the sample (observations) and \(\xi_0\) is the value of the mean specified by the null hypothesis. In this case under \(H_0\) the distribution of \(t'\) is Student's -t distribution with \(n_0-1\) degrees of freedom. The power function of the test is

\[
\beta(\xi) = 1 - P \left[ \frac{\xi_0 - \xi}{\sqrt{\sigma^2}} - \frac{t_{\alpha/2}}{\sqrt{\sigma^2}} < t_{n_0-1} < \frac{\xi_0 - \xi}{\sqrt{\sigma^2}} + \frac{t_{\alpha/2}}{\sqrt{\sigma^2}} \right] \tag{1.2.15}
\]

where \(\xi\) is the true mean.

The final sample-size \(N\) may assume any of the values \(\{n_0+1, n_0+2, \ldots\}\), and its distribution is given by

\[
P[N = n_0+1] = P[s^2/\sigma^2 \leq n_0+1]
= P \left[ \frac{(n_0-1)s^2}{\sigma^2} \leq \frac{(n_0-1)s^2}{\sigma^2} \right]
= P \left[ \chi^2_{n_0-1} \leq \frac{(n_0-1)s^2}{\sigma^2} \right] \tag{1.2.16}
\]

and, for all integers \(n > n_0+1\),

\[
P[N = n] = P[n < s^2/\sigma^2 + 1 \leq n+1]
= P \left[ \frac{(n-1)(n_0-1)s^2}{\sigma^2} < \frac{n(n_0-1)s^2}{\sigma^2} \right] \tag{1.2.17}
\]
The expected sample size, \( E(N) = \sum_n n P[N=n] \), satisfies the inequalities

\[
(n_0 + 1) \left[ P \left( \frac{\chi^2_{n_0-1}}{\sigma^2} < \frac{(n_0^2-1)z}{\sigma^2} \right) + \frac{\sigma^2}{z} P \left( \frac{\chi^2_{n_0+1}}{\sigma^2} > \frac{(n_0^2-1)z}{\sigma^2} \right) \right] \leq E(N) \quad (1.2.18)
\]

and

\[
E(N) \leq (n_0 + 1) \left[ P \left( \frac{\chi^2_{n_0-1}}{\sigma^2} < \frac{(n_0^2-1)z}{\sigma^2} \right) + \frac{\sigma^2}{z} P \left( \frac{\chi^2_{n_0+1}}{\sigma^2} > \frac{(n_0^2-1)z}{\sigma^2} \right) \right] + P \left[ \frac{\chi^2_{n_0-1}}{\sigma^2} > \frac{(n_0^2-1)z}{\sigma^2} \right], \quad (1.2.19)
\]

hence

\[
0 \leq \lim_{\sigma \to \infty} \{E(N) - \sigma^2/z\} \leq \lim_{\sigma \to \infty} \{E(N) - \sigma^2/z\} \leq 1 \quad (1.2.20)
\]

and we may use \( \sigma^2/z \) as an approximation to \( E(N) \) for large \( \sigma^2 \).

While procedure \( P_s \) provides a level \( \alpha \) test of Student's Hypothesis with power independent of \( \sigma^2 \), for practice Stein suggests a modified level \( \alpha \) test whose power function (while not independent of \( \sigma^2 \)) is uniformly higher than that of the test based on \( F' \); the modified test requires slightly smaller samples also. For example for Student's Hypothesis the procedure would be given as follows: instead of (1.2.8) define \( N \) by

\[
N = \max \{n_0, \lfloor s^2/z \rfloor + 1\}, \quad (1.2.21)
\]
and reject $H_0$ if and only if $|t''| > t_{n_0-1}^{\alpha/2}$. Below we denote statistics of the form $\sum_{j=1}^{N} a_j X_j$ by $\bar{X}$ and refer to them as generalized sample means. We also denote by

$$P_c(n_0; a; b, z)$$

(1.2.23)

a two-stage sampling procedure, indexed by $c$, which takes an initial sample of size $n_0$ and defines the final sample size $N$ by

$$N = \max\{n_0 + a, [b/z] + 1\}$$

(1.2.24)

Five years after Stein's 1945 paper, considerations of other problems plagued by unknown variances began to appear in the literature. The Behrens-Fisher problem (of testing the equality of the means of two independent normal populations when the variances are unknown) has been considered since 1929 by such scientists as W. U. Behrens, R. A. Fisher, J. Neyman, B.C. Welch, A. Aspin, M.S. Bartlett, H. Scheffe', and many others, with respect to single-sample solutions. Chapman (1950) developed a two-stage test for the Behrens-Fisher problem which had power independent of the unknown variances; his actual problem (slightly more general than the Behrens-Fisher) was: assume $\pi_1 \sim N(\mu_1, \sigma_1^2)$ and $\pi_2 \sim N(\mu_2, \sigma_2^2)$, define $k$ by $\mu_1 = k\mu_2$, and consider testing

$$H_0 : k = k_0.$$
(Thus if \( k_0 = 1 \) we have the Behrens-Fisher problem's hypothesis.) Chapman's test is based on the difference of two independent and identically distributed (i.i.d) Student's-t variates, and is a special case of our procedure \( \mathcal{P}_{E_1} \) given below in the case of the one-way layout.

Two-stage procedures have also been successfully constructed for problems of simultaneous estimation. Healy (1956) constructed simultaneous confidence intervals of fixed length and confidence coefficients \( 1 - \alpha \) for the means of \( k \) independent normal populations with a common unknown variance, for the \( k(k-1)/2 \) differences of means and for all possible normalized linear combinations of the \( k \) means. Hochberg (1975) generated simultaneous intervals for all \( k(k-1)/2 \) pairwise differences of means, for all contrasts, and for all linear combinations of the \( k \) means of \( k \) independent normal populations with unknown and unequal variances. His intervals are based on sample means, rather than generalized sample means, have random lengths which are less than a prespecified value \( L \) with probability one, and have overall confidence coefficient \( 1 - \alpha \).

Finally, for the related statistical problems of ranking and selection of \( k \) independent normal populations with unknown and unequal variances, Dudewicz and Dalal (1975) developed procedures based on generalized sample means which yield a probability of correct selection, denoted \( P[CS] \), which is at least \( P^* \) (prespecified) and independent of the unknown variances. They also suggested the possibility of such solutions for multiple-comparison problems under conditions of heteroscedasticity. Dudewicz (1972, page 80) suggested the application of two-stage procedures to the problems of ANOVA and MANOVA.
Rinott (1975) considered questions relating to sample means which had been raised by Dudewicz and Dalal, and gave new procedures based on sample means. (Details of the above works are given in the review of Section 1.3.)
1.3 Review of the Literature

We now review the important literature of heteroscedasticity, considering in detail those parts of the literature which relate most directly to our work; Stein's 1945 paper was discussed in Section 1.2 and will therefore not be included here. The literature naturally falls into five categories: papers dealing with two-stage sampling and relating directly to Stein's 1945 paper, papers considering the Behrens-Fisher problem (which is a special case of the problem we consider); multiple-comparison papers which assume unequal variances; papers dealing specifically with the analysis of variance problem under heteroscedastic conditions, and the two-stage ranking and selection papers.

In Table 1.3.1 we have arranged the papers considered in the review by these five categories. The author and year is given for easy reference and the page they appear in the dissertation is also given.
**TABLE 1.3.1**

<table>
<thead>
<tr>
<th>Category</th>
<th>Papers</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Related Directly to Stein (1945)</td>
<td>Dantzig (1940)</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Stein (1945)</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Chatterjee (1959a)</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Chatterjee (1969b)</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Ruben (1961)</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Ruben (1961a,b,c)</td>
<td>19,20,21</td>
</tr>
<tr>
<td></td>
<td>Seelbinder (1953)</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Moshman (1958)</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Wormleighton (1960)</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Weiss (1955)</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Bhattacharyee (1965)</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Volodin (1973)</td>
<td>33</td>
</tr>
<tr>
<td>Related to Behrens-Fisher Problems</td>
<td>Welch (1937)</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Welch (1947)</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Scheffé (1943)</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>Gipps and Smart (1972)</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Scheffé (1970)</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Chapman (1950)</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Ghosh (1975a)</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Koopman and Qualls (1971)</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Bennett (1971)</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>James (1954)</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Yao (1965)</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Subrahmanian and Subrahmanian (1973)</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Srivastava (1970)</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Ghosh (1971)</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Dudewicz (1972)</td>
<td>44</td>
</tr>
<tr>
<td>Simultaneous Inference</td>
<td>Healy (1956)</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Ghosh and Sen (1971)</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>Chatterjee (1962)</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>Spjotvoll (1972)</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>Hochberg (1975)</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Dudewicz, Ramberg, and Chen (1975)</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>Andrews and Arnold (1974)</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Dudewicz and Ramberg (1972)</td>
<td>65</td>
</tr>
<tr>
<td>Category</td>
<td>Papers</td>
<td>Pages</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Heteroscedastic ANOVA</td>
<td>James (1951)</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Welch (1951)</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>James (1954)</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Brown and Forsythe (1974a)</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Brown and Forsythe (1974b)</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Mazuy and Connor (1965)</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>Chakravarti (1965)</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Kulkarni (1973)</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Demskey (1975)</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Box and Hill (1974)</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Anderson (1963)</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>Bhargava (1971)</td>
<td>82</td>
</tr>
<tr>
<td>Ranking and Selection</td>
<td>Dudewicz (1971)</td>
<td>231</td>
</tr>
<tr>
<td></td>
<td>Dudewicz and Dalal (1975)</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Rinott (1975)</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>Tamhane (1975)</td>
<td>87</td>
</tr>
</tbody>
</table>

Ruben wrote a series of articles which considered the same questions as Stein but in a generalized and stronger form. The first of these, Ruben (1961) deals with conservative procedures for the joint interval estimation of the means, via two-stage sampling, of \( k \) independent normal populations with a common unknown variance. Confidence sets of predetermined volume and confidence coefficient are derived and shown to be conservative in nature. Similarly conservative tests for the general linear hypothesis are developed which have power functions independent of the unknown variance.

The two-stage sampling scheme \( P(n_0, 0; s^2, z) \) is used on each of the \( k \) population where \( s^2 \) is the pooled estimate of the unknown variance based on the first-stage observations. Two-stage sample means \( \{\bar{x}\}_i^k \) are used (rather than generalized sample means). Let \( N \) be the final sample-size taken from each population, \( N = \max[n_0, \lfloor s^2/z \rfloor + 1] \).
Let \( \mu = (\mu_1, \ldots, \mu_k) \) be the vector of means, \( \sigma^2 \) the unknown variance, 
\( \xi \) a random vector distributed as \( N_k(0, I) \) and \( R \) any given subset of 
Euclidean \( k \) space denoted by \( \mathbb{R}^k \). The exact probability distribution 
of the vector of sample means, \( \bar{X} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k) \), is given by

\[
P[\bar{X} \in R|\mu, \sigma] = \sum_{m=1}^{\infty} p_m(\sigma) Q_m(R;\mu, \sigma/\sqrt{m})
\]

where

\[
p_m(\sigma) = P[N = m|\sigma] \tag{1.3.2}
\]

and

\[
Q_m(R;\mu, \sigma/\sqrt{m}) = \int \cdots \int (2\pi)^{-k/2} e^{-\frac{1}{2} \xi^T \xi} d\xi. \tag{1.3.3}
\]

Ruben's approach is to consider the limiting distribution of \( \bar{X} \) as 
\( \sigma \to \infty \), which he shows may be expressed as

\[
P^*(R|\mu) = \int_0^\infty P^*(R;\mu, \sqrt{r_0}/u) dF_{r_0}(u) \tag{1.3.4}
\]

where \( F_{r_0}(u) \) is the cumulative distribution function (CDF) of a \( \chi^2_{r_0} \) random variable, \( r_0 = n_0 - 1 \), and

\[
Q^*(R;\mu, \sqrt{r_0}/u) = \int \cdots \int (2\pi)^{-k/2} e^{-\frac{1}{2} \xi^T \xi} d\xi. \tag{1.3.5}
\]

From this it follows that in the limit \( (\sigma \to \infty) \), \( (\bar{X} - \mu)/\sqrt{2} \) is distribu-
ted as \( (1/\sqrt{U/r_0})\xi \), where \( U \sim \chi^2_{r_0} \).
Tests and estimation procedures are based upon this limiting distribution (i.e. assuming \((\bar{x}_i - \mu) / \sqrt{z} \sim (1/\sqrt{U_0})_{i}^2\)), and it is shown that the true confidence coefficients for the confidence sets are greater than the confidence coefficients obtained from the limiting distribution. The true power curves for the tests and the power curves based on the limiting distribution are shown to cross with the true power greater than the limiting power for alternatives sufficiently far from the null hypothesis. This leads to the claim of conservativeness for the procedures and, since the limiting distribution for \(\bar{X}\) is the same as the distribution of Stein's generalized sample means, we see that Ruben's work is a stronger version of Stein's original solutions.

Ruben (1962a) generalizes Ruben (1961) to the case of unknown and unequal variances. The two-stage sampling scheme \(P(\eta_{i}, 0, s_{i}^{2}, z_{i})\) is used on the \(i\)th population, where \(s_{i}^{2}\) is the estimate of the variance \(\sigma_{i}^{2}\) of the \(i\)th population based on the first-stage sample. The generalization is straightforward and again the limiting distribution of the vector \(\bar{X} = (\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k)\) is considered, this time as all \(\sigma_{i} \to \infty\) independently and arbitrarily. The limiting distribution in this case is given by

\[
P^*(R/\mu) = P\left[\left(\mu_{i} + \sqrt{z_{i}} \, t_{1}\right), \ldots, \left(\mu_{k} + \sqrt{z_{k}} \, t_{k}\right) \in R\right] (1.3.6)
\]

where \(R\) is any fixed subset of \(\mathbb{R}^k\), \(\{t_{i}\}_{i=1}^k\) are independent Student's \(t\) variates with respective \(n_{0_i} - 1\) degrees of freedom and \(\mu_{i}\) is the mean of the \(i\)th population. Confidence interval estimation and testing is extensively covered for the case \(k = 2\), and the confidence intervals are shown to be conservative.
The limiting power curve and the true power curve for the test of the hypothesis $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 > \mu_2$ are shown to cross with the true power exceeding the limiting power if $\mu_1 - \mu_2$ is sufficiently greater than zero. For $k > 2$ similar results hold in estimation and testing for certain choices of the region $R$.

Efficiencies of the estimation and testing procedures relative to the fixed-sample case with known variances are considered, and it is shown that if the variance of the $i$th population exceeds $z_i n_{0i}$ then the efficiency of the two-stage procedure relative to the single-sample procedure is less than one but converges rapidly to one as $n_{0i} \to \infty$. (The efficiency of the two-stage procedure relative to the single-sample procedure is defined to be the ratio of the single-stage sample size needed to have confidence coefficient $1-\alpha$ for a confidence interval on the difference $\mu_1 - \mu_2$ when $\sigma_1^2$ and $\sigma_2^2$ are known, to the expected number of observations needed to achieve the same results for the two-stage procedure).

Ruben (1962b) used his two-stage sampling scheme to obtain confidence intervals of predetermined length and confidence coefficient for the mean of a stratified population with $r$ normal components. The mean in this case is defined to be the parameter $\mu = \sum_{i=1}^{r} p_i \mu_i$, where $\mu_i$ is the mean of the $i$th stratum (which has variance $\sigma_i^2$) and $p_i$ is the weight for the $i$th stratum.

Ruben (1962c) considers $k$ independent normal populations such that the $i$th population $\pi_i \sim N(\mu_i, \sigma_i^2)$. The two-stage sampling procedure $\mathcal{P}(n_{0i}, \sigma_i^2, z_i)$ is used on population $\pi_i$ and confidence regions, which are ovaloids with fixed semi-axes and confidence
coefficient, are obtained for the joint estimation of the vector of means $\mu = (\mu_1, \ldots, \mu_k)$. In this case the statistic proposed is

$$Y = \prod_{i=1}^{k} \left(1 + \frac{1}{(n_0 - 1)z_i} \left(\bar{X}_i - \mu_i\right)^2\right)^{-n_0_i/2}.$$  \hspace{1cm} (1.3.7)

As before the limiting distribution of $Y$ is considered as all $\sigma_i \to \infty$ independently and arbitrarily, and it is shown that in the limit $Y$ is distributed as

$$\prod_{i=1}^{k} \left(1 + \frac{t_i^2}{n_0 - 1}\right)^{-n_0_i/2}.$$  \hspace{1cm} (1.3.8)

where $\{t_i\}_{i=1}^{k}$ are independent, $t_i \sim t_{n_0 - 1}$. The distribution of (1.3.8) is unknown and Ruben suggests using $-2 \ln Y$ which is asymptotically (as all $n_0_i \to \infty$) distributed as $\chi^2_k$.

The expected total sample size $E(N)$ of Stein's two-stage procedure is of considerable interest, and depends upon the choice of the initial sample-size $n_0$. Seelbinder (1953) considered several methods of choosing $n_0$ for the case of estimating a single population mean $\mu$ with an allowable discrepancy $d$.

Consider the sampling procedure $P(n_0, 0; s^2, d^2/4t^2)$ where $s^2$ is the estimate of the unknown variance based on the first-stage sample and $t$ is the upper $a/2$ percent point of the Student's-t distribution with $n_0 - 1$ degrees of freedom. Set $c = d/\sigma$ and let $N$ be the final total sample size. Then Seelbinder first shows that

$$E(N) = \left[n_0 - t^2/c^2\right] F(\nu_0^2) + t^2/c^2[1 + K]$$  \hspace{1cm} (1.3.9)

where
$\chi_0^2 = (n_0-1)c^2/t^2$,

$F(\cdot)$ is the CDF of a chi-square distribution with $n_0-1$ degrees of freedom and

$$K = \frac{(\chi_0^2/2)^{(n_0-1)/2}}{(n_0-1)/2 \, \Gamma(n_0-1/2) \, e^{\chi_0^2/2}}.$$ 

Seelbinder obtains an exact expression for $N$ because he assumes $N = \max(n_0+1, \frac{4s^2t^2}{d^2})$ neglecting the small discrepancy introduced by the fractional nature of $N$ in contrast to the bounds obtained for $E(N)$ by Stein.

Next a normal approximation for $E(N)$ is given and numerical results indicate that for $n_0 > 10$ a reasonable, though not a precise, approximation is given by

$$E(N) = [n_0-t^2/c^2] \, \phi(L) + \frac{t^2}{c^2} \left[ 1 + \frac{L + 2\sqrt{2n_0-3}}{2(n_0-1) \, e^{L^2/2} \, \sqrt{2\pi}} \right]$$

(1.3.10)

where

$$L = c/t \, \sqrt{2(n_0^2-1)} - \sqrt{2n_0-3}.$$ 

Tables evaluating the validity and describing the use of the approximation are provided.

Considering $n_0$, $\alpha$ and $t$ as fixed the limiting behavior of $E(N)$ is studied as $c^2 \to 0$ and also as $c^2 \to \infty$. It is shown that

$$\lim_{c^2 \to 0} E(N) = n_0 \quad \text{and} \quad \lim_{c^2 \to \infty} \left( E(N) - \frac{t^2}{c^2} \right) = 0.$$ 

For cases when it can
be assumed that $\sigma$ lies in a known interval (say $\sigma_1 \leq \sigma \leq \sigma_2$) and $d$ is fixed, a discussion is presented on the choice of $n_0$ to minimize the maximum "loss of observations" $D = E(N/n_0) - E^*(N)$ where $E(N/n_0)$ is the expected sample-size for fixed $c$ and initial sample-size $n_0$ and $E^*(N)$ is the total sample-size needed for fixed $c$ and $\sigma$ known. A numerical example is provided illustrating the method for choosing $n_0$ in this case.

When nothing is assumed about $\sigma$ Seelbinder considers an upper limit on $n_0$ based on numerical studies of $E(N/n_0)$ for various values of $n_0$ and $c$. The numerical results indicate that an initial sample size of about 250 for $c < .1$ will yield an expected total sample size close to the sample size one would need if $\sigma$ were known.

This figure of 250 is extremely high because Seelbinder's analysis is very conservative (i.e. minimax). The 250 is practical if $c$ is small, however, Seelbinder's own tables show that for $c \geq .30$ (i.e. $d \geq .3\sigma$) and $\alpha = .10, .05, .02, .01$ an $n_0$ between 10 and 20 yields a very reasonable expected sample-size. For example: if $\alpha = .05, c = .5$, and $n_0 = 10$, the expected sample-size would be 21. The figure proposed by Seelbinder as an upper limit is thus protecting the experimenter only when $\sigma$ is extremely large relative to the precision desired in the estimate, while for almost all situations Seelbinder's tables show that a reasonable choice for $n_0$ is between 5 and 30.

Typical values of $E(N)$ are tabulated for various values of $n_0$ and $c$. A portion of the table is given below.
**EXPECTED SAMPLE SIZES (α = .05)**

<table>
<thead>
<tr>
<th>(c)</th>
<th>.1</th>
<th>.2</th>
<th>.3</th>
<th>.4</th>
<th>.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n_0)</td>
<td>60</td>
<td>400</td>
<td>100</td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>416</td>
<td>104</td>
<td>46</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>435</td>
<td>109</td>
<td>48</td>
<td>28</td>
</tr>
</tbody>
</table>

Moshman (1958) considered the same problem as Seelbinder (i.e. the choice of the initial sample size \(n_0\)) and gave a rule-of-thumb (based on the expected sample size and the upper 100th percentile of the distribution of \(N\)) as a guide to the selection of \(n_0\).

Considering the sampling procedure \(P(n_0, 0; s^2, z)\) applied to a single population where \(s^2\) is the estimate of \(\sigma^2\) based on the first \(n_0\) observations, letting \(\lambda = z/\sigma^2\), Moshman expressed the expected sample size by

\[
E(N) = n_0 P[X_{n_0}^2 - 1 \leq (n_0 - 1)n_0 \lambda] + \frac{1}{\lambda} P[X_{n_0+1}^2 > (n_0 - 1)n_0 \lambda]
\]

\[
+ \Theta_1 P[X_{n_0+1}^2 > (n_0 - 1)n_0 \lambda]
\]

(1.3.11)

using \(\Theta_1\) to obtain an exact expression for \(E(N)\) which does not assume a fractional form for \(N\) as Seelbinder (1.3.9), and
\[ \text{Var}(N) = \frac{n_0^2 \cdot \text{Pr}[X^2_{n_0-1} < (n_0-1)n_0 \lambda]}{(n_0-1)\lambda} + \frac{n_{0}^{+1}}{(n_0-1)\lambda} \cdot \text{Pr}[X^2_{n_0+3} > (n_0-1)n_0 \lambda] \\
+ \frac{2\varphi}{\lambda} \cdot \text{Pr}[X^2_{n_0+1} > (n_0-1)n_0 \lambda] + \theta \cdot \text{Pr}[X^2_{n_0-1} > (n_0-1)n_0 \lambda] - (E(N))^2 \]

\[ \text{(1.3.12)} \]

where $0 \leq \theta_i \leq 1$ ($i = 1, 2, 3$).

An efficient choice of $n_0$ would yield a small expected sample size and reduce the probability of large values of $N$. However the value of $n_0$ which minimizes $E(N)$ does not in general minimize $\text{Var}(N)$ and thus an optimal choice of $n_0$ is not obvious. Moshman suggests the following combination of $E(N)$ and percentage points of the distribution of $N$ as a guide to the choice of $n_0$. Define $N_p$ to be the smallest integer such that $\text{Pr}[N \leq N_p] = \sum_{m=n_0}^{N_p} \text{Pr}[N = m] \geq p$, which is equivalent to choosing $N_p$ so that

\[ \int_0^{(n_0-1)N_p \lambda} f_{n_0-1}(x) \, dx \geq p \]

\[ \text{(1.3.13)} \]

where $f_{n_0-1}(x)$ is the density function of a $\chi^2_{n_0-1}$ random variable.

Let $E(N/n_0^\star)$ be the expected value of $N$, and $N_p(n_0^\star)$ the 100th percentile of $N$, given $n_0 = n_0^\star$. Set

\[ \text{Pr}(n_0^\star) = \int_0^{(n_0-1)\lambda \cdot E(N/n_0^\star)} f_{n_0-1}(x) \, dx = \text{Pr}[N \leq E(N/n_0^\star)] \]

which is the proportion of time $N$ will not exceed $E(N/n_0^\star)$, and let $n_0^{**}$ be the value of $n_0$ which minimizes $E(N)$. Then consider
\[ \Psi(n_0) = (1-p)(N_p(n_0^**) - N_p(n_0)) - (1-P(n_0^**))(E(N/n_0) - E(N/n_0^*)) . \] (1.3.14)

\[ \Psi(n_0) \] weights the expected changes in \( E(N) \) and \( N_p \) by the probability that \( N \) exceeds those values. Moshman suggests choosing \( n_0 \) to maximize \( \Psi(n_0) \). That this is reasonable becomes clearer by writing

\[ -\Psi(n_0) = (1-p) N_p(n_0) + (1-P(n_0^**)) E(N/n_0) - K \] (1.3.15)

where

\[ K = (1-p) N_p(n_0^**) + (1 - P(n_0^**)) E(N/n_0^**) . \]

Now we want to choose \( n_0 \) to make \( N_p(n_0) \) and \( E(N/n_0) \) small, which is achieved by minimizing \( -\Psi(n_0) \) (or, by maximizing \( \Psi(n_0) \)).

An example with \( p = .95 \) and \( \lambda = .1 \) shows that this rule-of-thumb suggests a value of \( n_0 = 6 \).

Chatterjee (1959a) provided a two-stage procedure for testing the hypothesis that the mean vector \( \mu = (\mu_1, \ldots, \mu_p)' \) of a \( p \)-dimensional (multivariate) normal distribution is equal to \( \mu_0 \), a fixed known vector, such that the power function is independent of the unknown covariance matrix \( \Sigma = (\sigma_{ij}) \). Without loss of generality we assume that \( \mu_0 = 0 \). The procedure is: arbitrarily select \( z > 0 \), an integer \( n_0 > p \), and a \( p \times p \) positive-definite matrix \( \Sigma_{rs} \). Take \( n_0 \) initial observations \( X_1, \ldots, X_{n_0} \) where \( X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})' \) (i = 1, 2, ..., \( n_0 \)) and compute

\[ \bar{X}_i = \frac{1}{n_0} \sum_{\ell=1}^{n_0} x_{i\ell}, \quad s_{ij} = \sum_{\ell=1}^{n_0} (X_{i\ell} - \bar{X}_i)(\bar{X}_j - \bar{X}_j), \]

\[ s_{ij} = \frac{1}{(n_0-1)} S_{ij} \quad i, j = 1, 2, \ldots, p . \] (1.3.16)
Define the positive integer $N$ by
\[ N = \max\{n_0 + p^2, \left[ z \sum_{i,j=1}^p \alpha_{ij}s_{ij} \right] + 1 \}, \quad (1.3.17) \]
and select $p$ $(p \times N)$ matrices
\[
A_r = \begin{bmatrix}
a_{r11} & \cdots & a_{r1N} \\
\vdots & \ddots & \vdots \\
a_{rp1} & \cdots & a_{rPN}
\end{bmatrix} \quad (r = 1, 2, \ldots, p)
\]
in such a way that:

1) $a_{r11} = \cdots = a_{r1n_0}$

2) $A_r\eta = \epsilon_r$ where $\eta$ is the $N \times 1$ vector $(1, 1, \ldots, 1)^T$ and $\epsilon_r$ is the $p \times 1$ vector whose $r$th element is 1 and all other elements are 0; and

3) $AA' = \frac{1}{z} (\alpha^{rs}) \otimes (s^{ij})$, where $A = (A_1, A_2', \ldots, A_p')$ and $(\alpha^{rs}) \otimes (s^{ij})$ is the direct product of matrices (see Section 1.2) $(r, i = 1, 2, \ldots, p)$.

(It is shown that the conditions the final sample size must satisfy to insure the existence of such matrices is (1.3.17).) Next take $N-n_0$ additional observations $X_{n_0+1}', \ldots, X_N'$ and compute
\[
\xi_r = \sum_{i=1}^p \sum_{\ell=1}^N a_{r\ell i} X_{i\ell} \quad (r = 1, 2, \ldots, p). \quad (1.3.18)
\]

For fixed $(s_{ij})$, $(\xi_1', \ldots, \xi_p') \sim N_p (\mu, V)$ where $V = \frac{1}{z} (\alpha^{rs})$ and $L = \sum_{i,j}^p \sigma_{ij}s_{ij}$. The test statistic proposed by Chatterjee is
\[ U = z \sum_{r,s=1}^{p} \alpha_{rs} u^r v^s, \quad (1.3.19) \]

which is shown to have density

\[
f(u) = \sum_{t=0}^{\infty} \frac{1}{t!} \left\{ \frac{\Delta}{2(n_0-1)} \right\}^t \frac{1}{2} \frac{1}{\Gamma(b+2t)} \left\{ \frac{u}{n_0-1} \right\}^{(p+2t)/2-2} \\
\cdot \prod_{0 < \lambda_i < \cdots < \lambda_p} \exp \left\{ -\frac{u + \Delta}{2(n_0-1)} \sum_{i=1}^{p} 1/\lambda_i \right\} \left( \sum_{i=1}^{p} \frac{1}{\lambda_i} \right)^{-(p+4t)/2} \\
\cdot p(\lambda_1, \ldots, \lambda_p) \, d\lambda_1 \cdots d\lambda_p \quad (1.3.20)\]

where \( 0 < u < \infty, (\lambda_1, \ldots, \lambda_p) \) are the roots of the equation

\[ |(s_{ij}) - \lambda(s_{ij})| = 0, \]

\[ p(\lambda_1, \ldots, \lambda_p) = \exp \left\{ -\sum_{i=1}^{p} \lambda_i \right\} \prod_{i=1}^{p} \lambda_i \prod_{i=1}^{p} (\lambda_i - \lambda_j), \]

\[ 0 < \lambda_p < \cdots < \lambda_1, \quad (1.3.21) \]

is the joint density for \((\lambda_1, \ldots, \lambda_p)\), \( \Delta \) is the noncentrality parameter for the distribution of \( U \), and

\[ \Delta = z \sum_{r,s=1}^{p} \alpha_{rs} u^r v^s. \]

If \( H_0 \) is true, then \( \Delta = 0 \) and \( U \) is said to follow a central distribution as given by \((1.3.20)\) with \( \Delta = 0 \).

The power function is determined by the noncentral distribution of \( U \), and depends only upon \( \Delta \) and is shown to be monotonically
increasing in $\Delta$. Since $\Delta$ is independent of the $\sigma_{ij}$'s so is the power function, which is hence controllable through $z$.

The test is carried out by rejecting $H_0$ if and only if $U > C_\alpha$ where $C_\alpha$ is the upper $\alpha$th percent point of the central distribution of $U$.

Ellipsoidal confidence regions with predetermined shape and confidence coefficient are constructed, and are of the form

$$\sum_{i,j} n_{rs} (\mu_{ij} - \xi_{rs})(\mu_{ij} - \xi_{rs}) \leq C^2$$  \hspace{1cm} (1.3.22)

where $C^2$ and $(\alpha_{rs})$ are prespecified. The choice of $(\alpha_{rs})$ is made depending upon the relative importance of deviations in mean of the different variables. Often these deviations are equally important, in which case we take $(\alpha_{rs})$ to be the $p \times p$ identity matrix.

Chatterjee (1959b) considers the same questions as Chatterjee (1959a) but derives a uniformly more powerful test and a better confidence region.

The sampling procedure is similar to that in Chatterjee (1959a): let $n_0$ be an initial sample size, let $(\alpha_{rs})$ and $(s_{ij})$ be defined as in Chatterjee (1959a), and let $\lambda_m$ be the maximum root of the equation

$$|\alpha_{rs} - \lambda(s_{ij})| = 0. \hspace{1cm} (1.3.23)$$

Define

$$N = \max \left\{ n_0, \left[ \frac{T^2_{1-\alpha} \lambda_m}{C^2} \right] + 1 \right\}$$

where $T^2_{1-\alpha}$ is the upper $\alpha$th percent point of the central Hotelling's
$T^2$-distribution. Take $N-n_0$ additional observations and let 
$(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p)$ be the pooled sample mean vector. The confidence 
region is then defined by

$$\sum_{i,j} s^{ij}(\mu_i - \bar{X}_i)(\mu_j - \bar{X}_j) \leq \frac{T^2_{I-\alpha}},$$

(1.3.24)

and it is shown that this confidence region is smaller, and requires
less sampling, than the confidence region given by (1.3.22).

A uniformly more powerful test than that based on (1.3.19)
is also constructed. In this case $n_0$ initial observations are taken 
and $N$ is defined by

$$N = \max\{n_0 + p^2 - 1, [z \sum_{i,j} \alpha_{ij}s_{ij}] + 1\}.$$

(1.3.25)

Again $p \times (p \times N)$ matrices are selected such that if

$$A_r = \begin{bmatrix}
a_{r11} & \cdots & a_{r1N} \\
& \vdots & \\
a_{rpr} & \cdots & a_{rpN}
\end{bmatrix}
$$

then the elements $a_{r_{ij}}$ satisfy

1) $a_{r_{il}} = \ldots = a_{r_{in_0}}$, \hspace{1cm} r, i = 1, 2, \ldots, p

2) $A_r \eta = \epsilon_r$, 

\hspace{1cm}
where \( \eta \) and \( \epsilon_r \) are defined as in Chatterjee (1959a) and

\[
3) \quad A' = \frac{1}{n} \phi(\alpha^{rs}) \otimes (s^{ij}),
\]

and \( \phi = \sum_{i,j} \alpha_{ij} s_{ij} \). Then let

\[
\xi_r = \sum_{i=1}^{p} \sum_{l=1}^{N} a_{r_l} x_{il} \quad (1.3.26)
\]

and base the test of \( H_0 : \mu = 0 \) on the statistic

\[
U' = \frac{1}{\phi} \sum_{r,s=1}^{P} \alpha_{rs} \xi_r \xi_s \quad (1.3.27)
\]

and reject \( H_0 \) if and only if \( U' \geq u_{1-\alpha} \) where \( u_{1-\alpha} \) is the upper \( \alpha \)th percent point of the central \( U \) distribution given in (1.3.20). It is shown that the test based on \( U' \) is uniformly more powerful than the test based on \( U \) and requires less sampling (although its power function does depend upon \( (\sigma_{ij}) \)).

Weiss (1955) considered the problem of finding a confidence interval of fixed length \( L \) and confidence coefficient at least \( \beta \) for the mean \( \mu \) of a normal distribution with unknown variance \( \sigma^2 \). Two-sample plans, the size of the second stage depending upon the observations in the first stage, are considered. The study is limited to those two-stage schemes which increase the center of the interval by \( c \) units if each observation is increased by \( c \) units and for which the size of the second sample depends only upon the differences among observations of the first stage. Under these conditions it is shown that the optimum confidence interval of length \( L \) (i.e. largest confidence coefficient uniformly in \( \sigma \)) is the interval centered at \( \bar{x} \), the mean of all of the observations.
Weiss also shows that two-sample procedures which make the size of the second-stage sample a nondecreasing function of the first-stage sample variance are optimal in the following sense: if \( G(m, \sigma, R') = P[N=m | \sigma] \) using sampling procedure \( R' \) and \( N \) is the size of the second stage sample and if \( \lim_{m \to \infty} G(m, \sigma, R') = 1 \) for all \( \sigma \), then there exists a sampling rule \( R \) based on the sample variance of the first-stage observations such that

\[
G(m, \sigma, R') \leq G(m, \sigma, R) \quad (1.3.28)
\]

for all \( m > m(\sigma) \) where \( m(\sigma) \) is a finite integer.

Bhattacharjee (1965) studied the effect of non-normality on Stein's \( t \) statistic by deriving its distribution for non-normal populations represented by the first four terms of an Edgeworth series. The study shows that Stein's \( t \) is more sensitive to non-normality of the underlying population than Student's-\( t \). If the underlying population is not normal the power function of Stein's test for Student's Hypothesis is shown to depend upon the variance \( \sigma^2 \).

Departure from normality is measured by a population's skewness \( (\tau_3) \) and kurtosis \( (\tau_4) \). If \( t_{n_0-1}^{\alpha} \) denotes Stein's \( t \)-statistic the true \( P[t_{n_0-1} < t_{n_0-1}^{\alpha}] \) is tabled for various values of \( \tau_3, \tau_4 \) and \( a = z/\sigma^2 \) for \( \alpha = .025 \) and .005 which indicates a definite dependence on \( \sigma^2 \) through \( a \). For example if \( n_0 = 10, \alpha = .025, \tau_3 = -1.0, \tau_4 = 0 \) and \( a = 1.0 \) \( P[t_{n_0-1} < t_{n_0-1}^{.025}] = .0133 \) while if \( a = .001 \) \( P[t_{n_0-1} < t_{n_0-1}^{.025}] = .0240 \). Values of the power function of Stein's two-sided \( t \)-test at the 5\% level with \( n_0 = 5 \) were also tabled for various
values of \( \tau_3, \tau_4, a \) and alternatives \( \delta = |\mu - \mu_0| \). Here for example when \( n_0 = 5, \tau_3 = -.6, \tau_4 = 2, \delta = 1.0 \) and \( a = 1 \). The power is .0753 while if \( a = .001 \) the power is .1114. Other tables and graphs indicating the sensitivity of \( t_{n_0-1} \) to non-normality and the effect of \( \sigma^2 \) on power are presented.

Volodin (1973) developed a procedure which yields a test of the hypothesis of equality of means in a one-way layout with equal but unknown variances which has a prespecified level \( \alpha \) and power \( \beta \).

Let \( \pi_1, \ldots, \pi_k \) be the \( k \) populations under consideration, so that \( \pi_i \sim N(\mu_i, \sigma^2) \). The null and alternative hypothesis are given by

\[
H_0: \sum_{i=1}^{k} (\mu_i - \mu)^2 = 0
\]

\[
\bar{\mu} = \frac{1}{k} \sum_{i=1}^{k} \mu_i
\]

and

\[
H_A: \sum_{i=1}^{k} (\mu_i - \mu)^2 \geq \Delta^2 \quad (\Delta \neq 0).
\]

An initial sample of size \( n_{0i} \) is taken from \( \pi_i \) \((i = 1, 2, \ldots, k)\) and a pooled estimate \( s^2 \) of \( \sigma^2 \) is obtained which is based on

\[
r_2 = \sum_{i=1}^{k} n_{0i} - k \text{ degrees of freedom}. \quad N_i - n_{0i} \text{ additional observations are taken on } \pi_i \text{ where } N_i = \max(n_{0i}, n) \text{ and } n = \lceil \max(2, zs^2/\Delta^2) \rceil + 1
\]

where \( z > 0 \) is a prespecified constant. The test is then based on the test statistic

\[
\mathcal{S}^2 = \sum_{i=1}^{k} N_i (\bar{X}_i - \bar{X})^2
\]

where \( \bar{X}_i \) is the overall sample mean and \( \bar{X} = \frac{\sum_{i=1}^{k} N_i \bar{X}_i}{\sum_{i=1}^{k} N_i} \).
The test procedure rejects $H_0$ if and only if

$$J^2 \geq \frac{z_2^2}{\Delta^2} C \quad (1.3.32)$$

and it is shown that if $C = (\Delta^2 z_p)/(1-p)$ where $p$ is the only root of the equation

$$\sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{r_2}{2} + i\right)}{\Gamma\left(\frac{r_2}{2}\right)i!} \frac{r_2/2}{(1-p)^i} F_{k-1+2i, r_2 + 2i} (p \alpha) = \beta, \quad (1.3.33)$$

where

$$F_{m,n}(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \int_0^x \frac{t^{m/2-1}}{(1+t)^{(m+n)/2}} \, dt$$

and

$$z = \frac{(k-1)}{r_2} F_{k-1, r_2}^\alpha$$

Then the test has level $\alpha$ and power $\beta$ regardless of the value of $\sigma$.

An approximation to the distribution of $J^2$ is given by the limiting distribution obtained as $r_2 = \sum n_i - k \to \infty$ with $k$ fixed. An example applying the procedure is also presented.

If an experimenter wants to estimate the mean, $\mu$, of a normal distribution and the variance is known, then the sample-size may be chosen to achieve any desired precision. In this case a satisfactory balance can be reached between the cost of experimentation and the precision of the estimate. However, when the variance is unknown and a single-sample procedure is used, the precision is no longer controllable. Stein's estimation procedure (see Stein (1945, p. 246)) allows
the experimenter to control the precision but the cost of experimentation is then controllable only in expectation at best. Wormleighton (1960) presents a generalization of Stein's procedure which "very nearly" controls both cost and precision. Note that this paper is closely related to the problem of transforming a one-stage sample problem into a two-stage sample problem considered in Section 2.9.

Let \( \pi \) be the given normal population with unknown mean \( \mu \) and unknown variance \( \sigma^2 \). Take \( n_0 \) initial observations and let \( s^2 \) be the usual unbiased estimate of \( \sigma^2 \). For any value of \( s^2 \) one can plot the curve

\[
N_s(L) = \left[ \frac{t_{n_0-1}s}{L} \right]^2 \quad (0 < L < \infty)
\]

where \( 2L \) is the length of the confidence interval. In advance, on each curve choose a single point \( (N_s(L),L) \) which defines a "cut" across the family of curves indexed by \( s^2 \). The choice of the point would be made on the balance desired between the cost of experimentation and the precision desired if \( \sigma^2 \) actually were equal to \( s^2 \). Once the first-stage observations are taken \( s^2 \) may be calculated, which determines a particular curve and because of the cut, a unique point, \( (N^*,L^*) \).

Take \( [N^*-n_0] + 1 \) additional observations, where \([x]\) denotes the greatest integer less than \( x \) and compute the overall sample mean \( \bar{X} \). Then, take as the interval estimate of \( \mu \) the interval \( \bar{X} \pm L^* \) which is shown to be a \((1-\alpha)\) percent confidence interval for \( \mu \). This procedure allows the experimenter to choose a balance between cost and precision after the first-stage sample is taken and \( \sigma^2 \) is estimated.
through the choice of a "cut". This is closely related to the problem of selecting \( L \) a priori and taking a single-sample and transforming it to a two-stage sample by adjusting the confidence coefficient to obtain a confidence interval of prescribed length. However it should be noted that neither sample size or the length of the confidence interval are truly controllable here.

We next review the relevant literature dealing with the Behrens-Fisher problem, which is a special case of the problems for which we seek solutions. (The whole of the literature of the Behrens-Fisher problem is voluminous and cannot be covered here; the classical results may be found in the works of W. V. Behrens (1929), R. A. Fisher (1936), M. S. Bartlett (1936), and B. L. Welch (1937).)

A generalization of the Behrens-Fisher problem was developed by Welch (1947). Let \( \eta \) be a population parameter which is to be estimated by the observed random variable \( Y \) which is normally distributed with variance \( \sigma_y^2 = \sum_{i=1}^{k} \tau_i \sigma_i^2 \). The \( \tau_i \) are assumed to be known positive constants and the \( \sigma_i^2 \) are unknown variances.

It is easy to see that the Behrens-Fisher problem is a special case of the above. Let \( \pi_1 \) and \( \pi_2 \) be two populations such that \( \pi_i \sim N(\mu_i, \sigma_i^2), i = 1,2, \) and define \( Y = \bar{X}_1 - \bar{X}_2 \), where \( \bar{X}_i \) is the mean of the sample from \( \pi_i \) based on \( n_i \) observations. Then \( Y \sim N(\mu_1 - \mu_2, 1/n_1 \sigma_1^2 + 1/n_2 \sigma_2^2) \) and we have an observed random variable \( Y \) with \( \eta = \mu_1 - \mu_2 \) and \( \sigma_Y^2 = 1/n_1 \sigma_1^2 + 1/n_2 \sigma_2^2 \) so \( \tau_i = 1/n_i \) (i = 1,2).

In general, the problem is to determine a quantity \( h \) (a function of the observations through \( s_i^2 \), the usual unbiased estimate of \( \sigma_i^2 \)
which has the property that

\[ P[(Y - \eta) < h(s_1^2, ..., s_k^2)] = \alpha \]  \hspace{1cm} (1.3.35)

A series solution is developed which expresses \( h(s_1^2, ..., s_k^2) \) to terms of the order \( 1/f_i \) (\( f_i \) being the degrees of freedom associated with \( s_i^2 \)):

\[
h(s_1^2, ..., s_k^2) = \xi \sqrt{\sum_{i=1}^{k} \tau_i s_i^2} \left[ 1 + \frac{(1+\xi^2)}{4} \frac{\sum_{i=1}^{k} \left( \frac{\tau_i s_i^2}{f_i} \right)}{\left( \sum_{i=1}^{k} \tau_i s_i^2 \right)^2} \right. \\
\left. - \frac{(1+\xi^2)}{2} \frac{\sum_{i=1}^{k} \left( \frac{\tau_i s_i^2}{f_i} \right)}{\left( \sum_{i=1}^{k} \tau_i s_i^2 \right)^2} + \frac{(3+5\xi^2+\xi^4)}{3} \frac{\sum_{i=1}^{k} \left( \frac{\tau_i s_i^2}{f_i} \right)}{\left( \sum_{i=1}^{k} \tau_i s_i^2 \right)^3} \right. \\
\left. - \frac{(15+32\xi^2+9\xi^4)}{32} \frac{\sum_{i=1}^{k} \left( \frac{\tau_i s_i^2}{f_i} \right)^2}{\left( \sum_{i=1}^{k} \tau_i s_i^2 \right)^2} \right]. \hspace{1cm} (1.3.36)
\]

where \( \xi \) is the upper \( 1-\alpha \) percent point of the distribution of a standard normal variate.

A non-series approximation is also presented which is often referred to as Welch's approximate degrees of freedom solution. In this case the statistic

\[
\frac{(Y - \eta)}{\sqrt{\tau_1 s_1^2 + \tau_2 s_2^2}} \hspace{1cm} (1.3.37)
\]

is assumed to be approximately distributed as \( t_\tau \) where
Scheffe' (1943) presented the following solution for the Behrens-Fisher problem based on the Student's-t distribution. Let $X_1, \ldots, X_m$ be a random sample from $\pi_1 \sim N(\mu, \sigma_1^2)$, and let $Y_1, \ldots, Y_n$ be a random sample from $\pi_2 \sim N(\eta, \sigma_2^2)$ ($m \leq n$). Define the "Scheffe" variables $d_i$ by

$$d_i = X_i - \sum_{j=1}^{n} c_{ij} Y_j \quad (i = 1, \ldots, m) \quad (1.3.39)$$

where the $\{c_{ij}\}$ are chosen so that

a) $\sum_{j=1}^{n} c_{ij} = 1 \quad i = 1, \ldots, m$

and

b) $\sum_{k=1}^{n} c_{ij} c_{jk} = c^2 \delta_{ij} \quad i = 1, \ldots, m, \quad j = 1, \ldots, n$

The $d_i$'s are i.i.d. normal random variables with mean $\delta = \mu - \eta$ and variance $\sigma^2 = \sigma_1^2 + c^2 \sigma_2^2$. We may therefore generate a confidence interval for $\delta$ of the form

$$\bar{d} - t_{m-1}^{\alpha/2}(Q/[m(m-1)])^{1/2} \leq \delta \leq \bar{d} + t_{m-1}^{\alpha/2}(Q/[m(m-1)])^{1/2} \quad (1.3.40)$$

where $\bar{d} = \frac{1}{m} \sum_{i=1}^{m} d_i$ and $Q = \sum_{i=1}^{m} (d_i - \bar{d})^2$. 

$$f = \frac{(\sum_{i=1}^{k} r_i s_i^2)^2 - 2(\sum_{i=1}^{k} (r_i s_i^4)/(f_i+2))}{(\sum_{i=1}^{k} (r_i s_i^4)/(f_i+2))} \quad (1.3.38)$$
The optimal choice of the \{c_{ij}\}, in the sense of minimizing \(\sigma_1^2 + c^2\sigma_2^2\) is that which minimizes \(c^2\). It is shown that the optimum choice is given by

\[
c_{ij} = \begin{cases} 
\delta_{ij} \frac{(m/n)^{1/2} - (mn)^{-1/2} + 1/n}{j \leq m} \\
\frac{1}{n}, \quad j > m,
\end{cases}
\] (1.3.41)

in which case \(d = \bar{X} - \bar{Y}\) where \(\bar{X} = \frac{1}{m} \sum_{i=1}^{m} X_i\) and \(\bar{Y} = \frac{1}{n} \sum_{j=1}^{n} Y_j\).

Comparisons of the expected length of the above confidence interval with the expected length of the corresponding confidence interval used when the ratio of \(\sigma_1^2\) and \(\sigma_2^2\) is known were made. For the above procedure the expected length (for the optimal choice of \(\{c_{ij}\}\)) is

\[
E(L) = t^{\alpha/2} C_{m-1} \frac{(\sigma_1^2 + (m/n)\sigma_2^2)^{1/2}}{\sqrt{m}} / m^{1/2},
\] (1.3.42)

while in the case of \(\theta^2 = \sigma_1^2/\sigma_2^2\) known

\[
E(L) = t^{\alpha/2} C_{m+n-2} \frac{(\sigma_1^2 + m/\sigma_2^2)^{1/2}}{\sqrt{m}} / m^{1/2}
\] (1.3.43)

where \(C_K = 8K^{-1/2} \Gamma\left(\frac{1}{2} K + \frac{1}{2}\right)/\Gamma\left(K/2\right)\). Hence the ratio \(R\) of (1.3.42) to (1.3.43) is

\[
R = (t^{\alpha/2} C_{m-1}) / (t^{\alpha/2} C_{m+n-2})
\] (1.3.44)

and \(\lim_{m \to \infty} R = 1\) since \(\lim_{m \to \infty} C_m = 2\) although the limit (as \(m \to \infty\)) of both (1.3.42) and (1.3.44) is 0.
A general proof is given which shows that there does not exist a linear form \( \ell \) and a quadratic form \( Q \) which is independent of \( \ell \) with \( h(\ell - (\mu - \eta))|f \sim \mathcal{N}(0,1) \) and \( Q|f^2 \sim \chi^2_{k-1} \) where \( f \) is some function of the parameters such that for some constant \( h \)

\[
h(\ell - (\mu - \eta))/[Q|k-1]]^{1/2}
\]

will have a t-distribution with more than \( m-1 \) degrees of freedom. This generalization is used to show that Scheffe's intervals have minimum expected length among all such intervals based on the t-distribution.

Gipps and Smart (1972) investigated the power of the preceding solution given by Scheffe and compared the power of his test to the power of the t-test which can be constructed when the ratio of the variances is known, and to the power of the t-test derived when \( n-m \) of the observations are discarded, which is the Neyman-Bartlett solution (see Dudewicz (1976), pages 309-310).

If \( k = \sigma_1^2/\sigma_2^2 \) is known then the test of the hypothesis \( H_0 : \mu = \eta \) is based on

\[
\frac{\bar{x} - \bar{y}}{\{S^2(m^{-1} + kn^{-1})\}^{1/2}} \sim t_{m+n-2}(\varphi) \tag{1.3.45}
\]

where

\[
(m+n-2)S^2 = (m-1)S_x^2 + (n-1)S_y^2 \tag{1.3.46}
\]

\( S_x^2 \) and \( S_y^2 \) are the usual unbiased estimate of \( \sigma_1^2 \) and \( \sigma_2^2 \) respectively, and

\[
\varphi = (\mu - \eta)/\sigma_1 \left( \mu^{-1} + kn^{-1} \right)^{1/2}. \tag{1.3.47}
\]

Applying Scheffe's procedure the test is based upon the statistic
\[ \frac{\bar{X} - \bar{Y}}{\left( \frac{S^2_{d/m}}{m} \right)^{1/2}} \sim t_{m-1}(\varphi) \quad (1.3.48) \]

where \( S^2_{d} = \sum_{i=1}^{m} (d_i - \bar{d})^2 / m - 1 \), \( d_i \) defined as in (1.3.39) for the optimal \( \{ c_{ij} \} \).

Finally discarding the last \( n-m \) observations on the second population and defining \( d_i = X_i - Y_i \) (\( i = 1, \ldots, m \)) we may test the hypothesis with the statistic

\[ \frac{\bar{d}}{(S^2_{d/m})^{1/2}} \sim t_{m-1}(\varphi^*) \quad (1.3.49) \]

where

\[ \bar{d} = \frac{1}{m} \sum_{i=1}^{m} d_i , \quad S^2_{d} = \sum_{i=1}^{m} (d_i - \bar{d})^2 / m - 1 , \]

and

\[ \varphi^* = \varphi \left( \frac{n+nk}{(n+nk)} \right)^{1/2} . \quad (1.3.50) \]

The power functions for a two-tailed test for the case \( k \) is known, Scheffe's solution, and the procedure which discards \( n-m \) values, are respectively,

\[ \beta_1 = P[ |t_{m+n-2}(\varphi)| > t_{m+n-2}^{\alpha/2} ] , \quad (1.3.51) \]

\[ \beta_2 = P[ |t_{m-1}(\varphi)| > t_{m-1}^{\alpha/2} ] , \quad (1.3.52) \]

and

\[ \beta_3 = P[ |t_{m-1}(\varphi^*)| > t_{m-1}^{\alpha/2} ] . \quad (1.3.53) \]

From this it is clear that (1.3.51) and (1.3.52) involve \( k \) only through \( \varphi \). Thus the performance of Scheffe's test relative to the test when \( k \) is known depends only upon the magnitude of \( \varphi \). (1.3.53)
involves \( k \) explicitly as (1.3.50) shows and the power of these tests are therefore sensitive to \( k \).

Numerical calculations of these power functions were made for various \( k, m, n \) and levels. The power loss (relative to the test when \( k \) is known) using Scheffe's test is shown to depend considerably on the ratio of the sample sizes. However, the authors point out this becomes less critical when the smallest sample size exceeds 8. The loss is usually less than 10\% and often less than 5\%.

Scheffe (1970) compared six solutions to the Behrens-Fisher problem and compared them on the basis of size of the corresponding confidence intervals and associated confidence levels. They are: the elementary d solution, Scheffe's 19\% solution, the Welch-Aspin solution, Welch's approximate t-solution, the Fisher-Behrens solution and the usual Student's-t solution (assuming equal variances).

On the basis of the comparisons made in the paper only three are judged practical for use by Scheffe. (Student's solution is impractical because it assumes equal variances and thus asymptotically the confidence level can be anywhere between 0 and 1, his own solution is discarded because it requires randomization and the Fisher-Behrens solution is not recommended because its intervals are always longer than those given by the Welch-Aspin procedure.) The remaining three intervals are all judged satisfactory as solutions to the Behrens-Fisher problem. The details of the six procedures can be found in the paper with a complete list of references.

The two-stage sampling procedure given by (1.2.23) has been successfully applied to the Behrens-Fisher problem, to provide tests
and confidence intervals which are independent of the unknown variances. Chapman (1950) let \( \{X_{ij}\} (i = 1, 2; j = 1, 2, \ldots) \) be independent random variables such that \( X_{ij} \sim N(\mu_i, \sigma_i^2) \). Chapman considered testing the hypothesis

\[
H_0 : \mu_1 = k_0 \mu_2 ,
\]

where \( k_0 \) is a prespecified constant, as follows: let \( \pi_i \sim N(\mu_i, \sigma_i^2) \) (\( i = 1, 2 \)) be the two populations and apply the sampling procedure \( \mathcal{D}(n_0, 1; s_i^2, z_i) \) to \( \pi_i \) where \( s_i^2 \) is the usual sample variance for population \( \pi_i \) and \( z_1/z_2 = k_0^2 \). It follows from Stein (1945) that if

\[
T_i = \sum_{j=1}^{N_i} a_{ij} x_{ij} / \sqrt{z_i} \quad \text{then} \quad T_i - \mu_i / \sqrt{z_i} \sim t_{n_0-1} ,
\]

hence

\[
S = (T_1 - \mu_1 / \sqrt{z_1}) - (T_2 - \mu_2 / \sqrt{z_2}) = (T_1 - T_2) + (\mu_2 / \sqrt{z_2} - \mu_1 / \sqrt{z_1})
\]

is distributed as the difference of two independent \( t_{n_0-1} \) random variables. Letting \( k = \mu_1 / \mu_2 \) and \( \Delta = \mu_1 / \sqrt{z_1} - \mu_2 / \sqrt{z_2} = \mu_2 / \sqrt{z_2} (k/k_0 - 1) \),

\[
T_1 - T_2 \sim S + \Delta
\]

and under \( H_0 : \mu_1 = k_0 \mu_2 \) we have \( \Delta = 0 \).

The test is carried out by rejecting \( H_0 \) if and only if

\[
|T_1 - T_2| > s_0^{\alpha/2}
\]

for the two-sided alternative, where \( s_0^{\alpha} \) is the upper \( \alpha \)th percent point of the distribution of \( S \).

The power of the test is given by
\[ \beta(\Delta) = P[S < -s_{0}^{\alpha/2} - \Delta] + P[S > s_{0}^{\alpha/2} - \Delta] \] (1.3.59)

which is independent of \( \sigma_{1}^{2} \) and \( \sigma_{2}^{2} \) and is controllable through the \( z_{i} \)'s.

Tables giving the percentage points for the distribution of \( S \) are presented for \( n_{0} = 2, 4, 6, 8, 10 \) and 12 and it is suggested that the normal approximation (i.e. \( n_{0} = \infty \)) be used for \( n_{0} > 12 \). In recent years these tables have been shown to be in error and new tables for this distribution may be found in Ghosh (1975b), Dudewicz (1972) and Dudewicz, Ramberg and Chen (1975).

Chapman also noted a convenient method for choosing the constants \( \{a_{j}\} \) in Stein's procedure. In addition to requiring \( a_{1} = \ldots = a_{n_{0}} = a \) also require \( a_{n_{0} + 1} = \ldots = a_{N} = b \) and (via

\[ n_{0}a + (N-n_{0})b = 1 \] (1.3.60)

and

\[ n_{0}a^{2} + (n-n_{0})b^{2} = z/s^{2} \] (1.3.61)

we have

\[ b = \frac{1}{N} \left( 1 + \sqrt{\frac{n_{0}(N-\bar{s}^{2})}{(N-n_{0})s^{2}}} \right) \] (1.3.62)

and

\[ a = \frac{1 - (N-n_{0})b}{n_{0}} \] (1.3.63)

Ghosh (1975a) used the two-stage sampling procedure (1.2.23) to obtain confidence intervals on the difference of means of two independent
normal populations with unknown variances. Two concepts (discussed below) are used to measure the efficiency of the confidence intervals, and comparisons are made between Ghosh's intervals and those of Chapman.

Let \( \pi_1 \) and \( \pi_2 \) be the two populations under consideration, \( \pi_i \sim N(\mu_i, \sigma_i^2) \) \( (i = 1, 2) \). Ghosh applies procedure \( P(n_0, 0, s_0^2, z) \) to each population, where \( s_0^2 \) is defined as follows: let \( (X_1, \ldots, X_{n_0}) \) and \( (Y_1, \ldots, Y_{n_0}) \) be the initial samples of size \( n_0 \) taken on \( \pi_1 \) and \( \pi_2 \) respectively, \( T_j = X_j - Y_j \), \( T = \frac{1}{n_0} \sum_{j=1}^{n_0} T_j \), and

\[
s_0^2 = \frac{1}{n_0 - 1} \sum_{j=1}^{n_0} (T_j - \overline{T})^2.
\]

The confidence interval for \( \delta = \mu_1 - \mu_2 \) is based on the overall sample means \( \overline{X} = \frac{1}{N} \sum_{j=1}^{N} X_j \) and \( \overline{Y} = \frac{1}{N} \sum_{j=1}^{N} Y_j \) where \( N \) is the final total sample-size. It can be shown as in Stein (1945) that

\[
\frac{(\overline{X} - \overline{Y} - \delta)N^{1/2}}{s_0} \sim t_{n_0 - 1}
\]

and the confidence interval is given by

\[
[\overline{X} - \bar{Y} - s_0 N^{-1/2} t_{n_0 - 1} \alpha/2, \overline{X} - \bar{Y} + s_0 N^{-1/2} t_{n_0 - 1} \alpha/2],
\]

which has length

\[
L' = 2s_0 N^{-1/2} t_{n_0 - 1} \alpha/2 < 2 \sqrt{z} t_{n_0 - 1} \alpha/2
\]

If we choose \( z = [4(t_{n_0 - 1}^2)^2]^{-1} \), then \( L' < L \), hence we have a confidence interval of controlled length whose confidence coefficient \( 1 - \alpha \) is independent of \( \sigma_1^2 \) and \( \sigma_2^2 \).

The merits of the above interval are judged on the basis of its Wolfowitz and Neyman accuracy.
**Definition (1.3.66):** Let $[U, U]$ be a confidence interval for a parameter $\theta$ such that $P[\theta \in [U, U]|\theta| \geq 1-\alpha$ and $U - U \leq L$. Then $Q(\theta_0) = P[\theta_0 \in [U, U]|\theta|]$ is the **Neyman Accuracy** of $[U, U]$.

$Q(\theta_0)$ measures the accuracy of the interval $[U, U]$ in excluding false values if $\theta_0 \neq \theta$.

**Definition (1.3.67):** For arbitrary constants $a > 0$ and $b > 0$ the **Wolfowitz accuracy** of the confidence interval $[U, U]$ is defined as

$$\omega(a, b) = aE([U-\theta]^2) + bE([U-\theta]^2),$$

which measures the closeness of the end points $U$ and $U$ to the true value $\theta$. Ghosh criticizes past comparisons of two-stage confidence interval procedures on the grounds that they only consider the average sample-size. He claims the Wolfowitz and Neyman accuracies should also be used.

Properties of Ghosh's procedure are considered and it is shown that for his interval (for any $\theta_0 \neq \theta$)

$$\lim_{n_0 \to \infty} Q(\theta_0) = 0, \quad (1.3.68)$$

$$\lim_{L \to 0} Q(\theta_0) = 0, \quad (1.3.69)$$

and (for any $a > 0$, $b > 0$)

$$\lim_{n_0 \to \infty} \omega(a, b) = 0, \quad (1.3.70)$$

$$\lim_{L \to 0} \omega(a, b) = 0. \quad (1.3.71)$$
Comparisons are made with the confidence intervals obtained by Chapman's procedure. Let $I_S = [\bar{U}_S, \bar{U}_S]$ denote Ghosh's interval and $E(N_S)$ the expected sample-size for his procedure. Similarly, let $I_C = [\bar{U}_C, \bar{U}_C]$ denote Chapman's interval and $E(N_C)$ the corresponding expected sample-size. The following distinctions were noted:

a) Ghosh's procedure avoids a second-stage sample with positive probability (Chapman's does not);

b) Ghosh claims the computational aspects of his procedure are simpler than Chapman's, although this is questionable in light of the randomization required by his procedure;

c) The length of $I_S$ is almost surely less than the length of $I_C$;

d) $\lim_{n_0 \to \infty} Q_S(\delta_0)/Q_C(\delta_0) = 0$ for $\delta_0 \neq \delta$ which is of questionable significance since both numerator and denominator converge to 0 as $n_0 \to \infty$.

e) $\frac{\omega_S(a,b)}{\omega_C(a,b)} = \left( \frac{n_0-1}{n_0-3} + \left( \frac{t_\alpha/2}{n_0-1} \right)^2 \right) \left( \frac{n_0-1}{n_0-3} \right) - r + \left( \frac{t_\alpha/2}{n_0-1} \right)^2 \right)^{-1}$

where $r = \lim_{L \to 0} E(N_S)/E(N_C)$ (r is tabled for several values of $n_0$ and $\alpha$); the values in the table indicate that if $\alpha \leq .05$ and $n_0 \geq 10$, $r > 1$ while for $\alpha \leq .10$ and $n_0 = 5$, $r \geq .942$. This indicates that for reasonable values of $n_0$ and $\alpha$ in most cases Chapman's procedure is superior in terms of expected sample size. We also note that it would be of interest to investigate $E(N_S)/E(N_C)$ in some non-limiting cases for various values of $\sigma_1^2$ and $\sigma_2^2$.

f) Although not mentioned Ghosh's method requires randomization and Chapman's does not (hence Scheffé (1970) would, presumably, also
have rejected Ghosh's procedure had it been available at that time). Also note that Ghosh's approach takes the same sample-size from each population (regardless of the relative sizes of the population variances), while Chapman's procedure adjusts for this fact.

Koopmans and Qualls (1971) present confidence intervals of fixed length and fixed confidence coefficient for the variance, \( \sigma^2 \), and the coefficient of variation, \( \sigma/\mu \), for a single normal population with mean \( \mu \) and variance \( \sigma^2 \). The sampling procedures are of the form (1.2.23). Similar intervals are also given for the difference in means of two independent normal populations with unknown and unequal variances and also for the ratio and difference of the variances. We shall describe in detail the confidence interval obtained on the difference of the two means.

The \( \pi_1 \sim N(\mu_1, \sigma_1^2) \) and \( \pi_2 \sim N(\mu_2, \sigma_2^2) \). Take an initial sample of size \( n_{0_i} \) on \( \pi_i \) and obtain upper 100 \( \sqrt{1 - \alpha/2} \) percent confidence limits for \( \sigma_1^2 \) on the basis of these first \( n_{0_i} \) observations. That is let \( \hat{\sigma}_1^2 = [(n_{0_i} - 1)s_1^2]/B_i \) where \( s_1^2 \) is the usual unbiased estimate of \( \sigma_1^2 \) and \( B_i \) satisfies

\[
P[\chi^2_{n_{0_i} - 1} \leq B_i] = 1 - \sqrt{1 - \alpha/2} \quad (i=1,2).
\]

Let \( k_\alpha \) be the solution of the equation

\[
\Phi(k_\alpha) = 1 - \alpha/4
\]
If $c_1$ and $c_2$ are the unit sampling costs for the two populations then the cost of the second sample is $c_1 n_1 + c_2 n_2$. The second-stage samples are selected so as to minimize the cost of the second sample subject to the restriction

$$k\alpha \left( \frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2} \right)^{1/2} \leq L$$

where $n_1$ is the size of the second stage sample from $\pi_1$. The explicit allocation of sample sizes then is

$$n_1 = \left( \frac{k_\alpha ^2 \sqrt{c_1} \hat{\sigma}_1 + \sqrt{c_2} \hat{\sigma}_2}{L^2} + 1 \right)$$

$$n_2 = \left( \frac{k_\alpha ^2 \sqrt{c_1} \hat{\sigma}_1 + \sqrt{c_2} \hat{\sigma}_2}{L^2} + 1 \right)$$

(1.3.72)

The $100(1-\alpha)$ percent confidence interval for $\mu_1 - \mu_2$ then becomes

$$\bar{x}_1 - \bar{x}_2 - L \leq \mu_1 - \mu_2 \leq \bar{x}_1 - \bar{x}_2 + L$$

(1.3.73)

where $\bar{x}_1$ is the sample mean based on all $n_0 + n_1$ observations.

It would be of interest to make a comparison of this procedure with Ghosh's and Chapman's along the lines of Ghosh (1975a). Such a comparison would be possible by taking $c_1 = c_2 = 0$.

Several single-sample procedures have been proposed to test the multivariate Behrens-Fisher hypothesis of equality of two mean vectors of multivariate normal distributions when the covariance matrices are unknown and unequal.
Bennett (1951) generalized Scheffé's (1943) approach to the multivariate Behrens-Fisher problem. Let \( \tau_1 \sim \mathcal{N}_p(\mu_1, \Sigma_1) \) and \( \tau_2 \sim \mathcal{N}_p(\mu_2, \Sigma_2) \) where \( \mu_1, \mu_2, \Sigma_1, \Sigma_2 \) are all unknown. It is desired to test

\[
H_0 : \mu_1 = \mu_2
\]

versus

\[
H_A : \mu_1 \neq \mu_2.
\]

Let \( X_{ij} \ (i = 1, 2; j = 1, \ldots, N_i) \) be a random sample from \( \tau_1 \). The method used to test (1.3.74) consists of defining the vector \( Y_\alpha \) \((\alpha = 1, \ldots, N_1)\) by

\[
Y_\alpha = X_{1\alpha} - \left( \frac{N_1}{N_2} \right)^{1/2} X_{2\alpha} + \frac{1}{(N_1 N_2)^{1/2}} \left\{ \sum_{\beta=1}^{N_1} X_{2\beta} - \left( \frac{N_1}{N_2} \right)^{1/2} \sum_{\gamma=1}^{N_2} X_{2\gamma} \right\} (N_1 \leq N_2)
\]

The \( \{Y_\alpha\}_{\alpha=1}^{N_1} \) then constitute a random sample from a multivariate normal distribution with mean \( \mu_1 - \mu_2 \) and unknown covariance matrix. Thus Hotelling's \( T^2 \) statistic, \( \bar{Y}'S^{-1}\bar{Y} \), where \( \bar{Y} = 1/N_1 \sum_{\alpha=1}^{N_1} Y_\alpha \) and \( S = [S_{ij}] \), \( S_{ij} = \sum_{\alpha=1}^{N_1} (Y_{i\alpha} - \bar{Y})(Y_{j\alpha} - \bar{Y}) \), may be computed for the \( Y \)'s and the hypothesis equivalent to (1.3.74), that the mean of the \( Y \)'s is zero, may be tested.

James (1954) considered the same set-up as in Bennett (1951) and defined

\[
S_1 = \sum_{j=1}^{N_1} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)' /[N_i(N_i-1)]
\]
where

\[ \bar{x}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij} . \]

He suggested basing the test of (1.3.74) on the statistic

\[ (\bar{x}_1 - \bar{x}_2)'S^{-1}(\bar{x}_1 - \bar{x}_2) \]

where

\[ S = S_1 + S_2 \]

and rejecting (1.3.74) if and only if

\[ (\bar{x}_1 - \bar{x}_2)'S^{-1}(\bar{x}_1 - \bar{x}_2) \geq h(S_1, S_2; \alpha) \]

where

\[ h(S_1, S_2; \alpha) = \chi^2_{p, \alpha} \left[ 1 + \frac{1}{2} \left\{ \frac{k_1}{2} + \frac{k_2 \chi^2}{p(p+2)} \right\} \right], \]

\[ k_1 = \sum_{i=1}^{2} \left\{ \text{tr}(S^{-1}_i) \right\}^2 / (N_i - 1) \]

and

\[ k_2 = k_1 + 2 \sum_{i=1}^{2} \left\{ \text{tr}(S^{-1}_i S^{-1}_i) / (N_i - 1) \right\}. \]

Yao (1965) considered the same procedure and test statistic as James (1954). However, he considered an approximate degree of freedom solution and took as the critical region

\[ (\bar{x}_1 - \bar{x}_2)'S^{-1}(\bar{x}_1 - \bar{x}_2) \geq \chi^2_{p, N_1} \]
where $T^2_{Q}(p,N_T)$ is the upper $\alpha$th percent point of Hotelling's $T^2$ statistic with $N_T$ degrees of freedom which are estimated by

$$\frac{1}{N_T} = \frac{2}{\sum_{i=1}^{N_1} \frac{1}{N_i}} \left\{ \frac{(\bar{x}_1 - \bar{x}_2)'S^{-1}_iS^{-1}_i(\bar{x}_1 - \bar{x}_2)}{(\bar{x}_1 - \bar{x}_2)'S^{-1}_i(\bar{x}_1 - \bar{x}_2)} \right\}^2$$

Here $N_T$ is a random variable with

$$\min(N_1,N_2) \leq N_T \leq N_1 + N_2.$$

Subrahmanian and Subrahmanian (1973) compared the solutions to the multivariate Behrens-Fisher problem given by Bennett (1951), James (1954), and Yao (1965). A Monte Carlo study was performed and a comparison of level and power of these tests was made.

The results indicate that Yao's and James' procedures do not yield correct $\alpha$-level tests (especially as the difference between the covariance matrices becomes more pronounced). Bennett's procedure of course gives an exact $\alpha$-level test.

The power of James' procedure is always higher than the other two procedures but the $\alpha$-level of James' test is inflated. Bennett's procedure has the poorest power (especially when the sample sizes are very unequal). Extensive tables of comparison of level and power between the three procedures are given.

Several authors have considered the nonparametric analog to the Behrens-Fisher problem.
Srivastava (1970) considered a sequential solution of the non-parametric analogue of the Behrens-Fisher problem. Let $X_{ij} \sim F_i(x)$ where $F_i$ is a continuous distribution function (possibly unknown) such that

$$E(X_{ij}) = \theta_i, \quad \text{Var}(X_{ij}) = \sigma_i^2 \quad (i = 1, 2; j = 1, 2, \ldots),$$

and let $\theta = \theta_1 - \theta_2$, then the following procedure will yield a confidence interval $I$ of length $2d$ such that $\lim_{d \to 0} P[\theta \in I] = \alpha$.

Define

$$\bar{X}_{1l} = l^{-1} \sum_{j=1}^{l} X_{1j}, \quad \bar{X}_{2m} = m^{-1} \sum_{j=1}^{m} X_{2j}$$

$$V_{1l}^2 = (l-1)^{-1} \sum_{j=1}^{l} (X_{1j} - \bar{X}_{1l})^2, \quad l > 1 \quad (1.3.79)$$

$$V_{2m}^2 = (m-1)^{-1} \sum_{j=1}^{m} (X_{2m} - \bar{X}_{2m})^2, \quad m > 1.$$

Let $\{a_n\}$ be any sequence of positive constants such that

$$\lim_{n \to \infty} a_n = a \quad (1.3.80)$$

where

$$P[-a < Y < a] = \alpha, \quad Y \sim N(0, 1)$$

The sequential procedure begins by taking $n_0 > 1$ observations on each population. We then sample one extra observation at a time and stop with the first $l + m \geq 2n$ such that
\[ V_{1\ell}(V_{1\ell} + V_{2m}) \leq \ell d^2/a_n^2, \quad V_{2m}(V_{1\ell} + V_{2m}) \leq m d^2/a_n^2 \]

where at any stage, after \( i \) observations on \( F_1 \) and \( j \) observations on \( F_2 \), the next observation is taken on \( F_1 \) or \( F_2 \) according as

\[ 1/j < V_{1\ell}/V_{2j} \quad \text{or} \quad i/j > V_{1\ell}/V_{2j} . \]

When the sampling is stopped construct the confidence interval

\[ I = (\bar{x}_{1\ell} - \bar{x}_{2m} - d, \bar{x}_{1\ell} - \bar{x}_{2m} + d) \]

This interval has the properties

\[ \lim_{d \to 0} P[\theta \in I] = \alpha \]

\[ \lim_{d \to 0} d^2\ell/a^2\sigma_1(\sigma_1 + \sigma_2) = 1 \quad \text{a.s.} \]

\[ \lim_{d \to 0} d^2m/a^2\sigma_2(\sigma_1 + \sigma_2) = 1 \quad \text{a.s.} \]

\[ \lim_{d \to 0} d^2(\ell+m)/a^2(\sigma_1 + \sigma_2)^2 = 1 \quad \text{a.s.} \]

The remainder of the paper considers the asymptotic efficiency (i.e. the ratio of \( E(N) \) to the sample size needed to meet the requirements when \( \sigma_1^2 \) and \( \sigma_2^2 \) are known, and the underlying population is normal) as \( d \to 0 \). The multivariate analogue is also considered.

Ghosh (1971) provided a procedure for obtaining a bounded length confidence interval for the difference of medians of two symmetric
(but otherwise unknown) distributions. The procedure is the two-sample analogue of the one-sample procedure proposed by Ghosh and Sen (1971) based on rank order statistics. It is shown that asymptotically (as the length of the confidence interval approaches zero) the confidence coefficient is $1 - \alpha$.

An extremely important area of statistical inference is that of multiple-comparisons or simultaneous inference. In general single-stage simultaneous inference procedures such as Tukey's or Scheffe's multiple-comparison techniques are plagued by the same problems as single-stage testing procedures, that is, they depend upon the unknown variance(s).

Stein used his two-stage sampling scheme to construct a confidence interval for the mean of a normal population which had fixed length and confidence coefficient independent of the unknown variance. The two-stage sampling procedure (1.2.23) has been successfully applied to simultaneous inference problems.

Healy (1956) provided simultaneous confidence intervals of fixed length and confidence coefficient $1 - \alpha$ for (1) all normalized linear functions of means, (2) all differences in means, and (3) the means of $k$ independent normal populations with a common unknown variance. He also constructed joint confidence intervals of fixed length and confidence coefficient $1 - \alpha$ for (4) all normalized linear functions of means, and (5) the mean vector of a $p$-dimensional multivariate normal distribution with unknown covariance matrix. We now consider these five problems and Healy's solution in detail.
Problem (1). Let $X_{i1}, X_{i2}, \ldots$ ($i = 1, 2, \ldots, k$, $k \geq 2$) be independent random variables, $X_{ij} \sim N(\mu_i, \sigma^2)$, $\mu_i$ and $\sigma^2$ being unknown. Let $n_0$ be a fixed positive integer, and take $n_0$ initial observations on the $i$th population and let $s^2$ be an unbiased estimate of $\sigma^2$ based on $m$ degrees of freedom which is independent of $\sum_{j=1}^{n_0} X_{ij}, X_{i(n_0+1)}, \ldots$ ($i = 1, \ldots, k$). Let $\ell$ be the prespecified length desired for the confidence intervals on the linear functions of the means $\sum_{i=1}^{k} \frac{c_{ir}}{c_i} \mu_i$, where $\sum_{i=1}^{k} c_{ir}^2 = 1$, and $1-\alpha$ be the prespecified joint confidence coefficient. (We denote by $\mathcal{L}$ the set of all such normalized linear functions of the means and the index $r$ specifies a particular element of $\mathcal{L}$). Apply the two-stage sampling scheme $\mathcal{P}(n_0; 0; s^2, z)$ to each population and choose $z$ so that $P[W_1 \leq \ell^2/4z] = 1-\alpha$, where $W_1/k \sim F_{k,m}$. Take $N-n_0$ additional observations in each population where $N$ is the total sample size for each population defined by $\mathcal{P}(n_0; 0; s^2, z)$ and estimate $\sum_{i=1}^{k} c_{ir} \mu_i$ by the interval

$$
\frac{1}{N} \sum_{i=1}^{k} c_{ir} \sum_{j=1}^{N} X_{ij} \pm \frac{\ell}{2} \quad (1.3.81)
$$

It is then shown that

$$
P\left[ \frac{1}{N} \sum_{i=1}^{k} c_{ir} \sum_{j=1}^{N} X_{ij} - \frac{\ell}{2} \leq \sum_{i=1}^{k} c_{ir} \mu_i \leq \frac{1}{N} \sum_{i=1}^{k} c_{ir} \sum_{j=1}^{N} X_{ij} + \frac{\ell}{2} \forall r \right] \geq 1-\alpha \quad (1.3.82)
$$
Problem (2). Consider the same set up as in Problem (1) except now we seek joint confidence intervals of length $\ell$ and overall confidence coefficient $1-\alpha$ on the $k(k-1)/2$ differences $\mu_i - \mu_j$, $1 \leq i < j \leq k$. Use the same sampling scheme as above but choose $z$ so that 

$$P[W_2 \leq \ell/2 \sqrt{z}] = 1-\alpha,$$

where $W_2$ follows a studentized range distribution with $k$ and $m$ degrees of freedom. Now estimate $\mu_i - \mu_j$ by the interval

$$\frac{1}{N} \sum_{\ell=1}^{N} (X_{i\ell} - X_{j\ell}) \pm \frac{\ell}{2}$$

(1.3.83)

and

$$P\left[ \frac{1}{N} \sum_{\ell=1}^{N} (X_{i\ell} - X_{j\ell}) - \frac{\ell}{2} \leq \mu_i - \mu_j \leq \frac{1}{N} \sum_{\ell=1}^{N} (X_{i\ell} - X_{j\ell}) \right] \text{ for all } i \neq j \geq 1-\alpha$$

(1.3.84)

Problem (3). In this problem we take $n_{0i}$ initial observations on the $i$th population, are interested in obtaining a confidence interval of length $\ell_i$ on $\mu_i$, and want the overall joint confidence coefficient to be at least $1-\alpha$.

Select $z_1, \ldots, z_k$ such that $P[W_1 \leq \ell_i/2 \sqrt{z_i}] = 1 - \alpha$ ($i = 1, \ldots, k$), where $W_1$ follows the studentized maximum modulus distribution with $k$ and $m$ degrees of freedom. Apply the two-stage sampling scheme $\mathcal{A}(n_{0i}; 0; s^2, z_i)$ to the $i$th population and let $N_i$ be the final total sample size for the $i$th population. Estimate $\mu_i$ by the interval
The arguments needed to justify the preceding probability statements are claimed to be similar to the distribution arguments in Stein (1945).

Problem (4). Problems 4 and 5 deal with multiple comparison procedures applied to the multivariate normal population \( N_k(\mu, \Sigma) \), \( \mu = (\mu_1, \ldots, \mu_k)' \), \( \Sigma = (\sigma_{ij}) \), where \( \mu \) and \( \Sigma \) are unknown. Let \( \ell > 0 \) and \( 0 < 1 - \alpha < 1 \) be prespecified and seek a system of simultaneous confidence intervals for all functions \( \sum_{i=1}^{k} c_{ir} \mu_i \) with \( \sum_{i=1}^{k} c_{ir}^2 = 1 \) with length \( \ell \) and joint confidence coefficient \( 1 - \alpha \). We again denote this set of functions by \( \mathcal{S} \) and \( r \) is an index for the set. The solution is based on the two-stage sampling scheme \( \mathcal{Y}(n_0, 0; \tau, z) \) where \( \tau \) is the largest eigenvalue of the usual unbiased estimate of \( \Sigma \) based on the first \( n_0 \) observations. Choose \( z \) so that

\[
P[W_4 \leq \ell^2/4z] = 1 - \alpha \tag{1.3.87}
\]

where

\[
(m-k+1)W_4/k \sim F_{k, m-k+1}.
\]

Now estimate \( \sum_{i=1}^{k} c_{ir} \mu_i \) by

\[
\frac{1}{N_i} \sum_{\ell=1}^{N_i} X_{i\ell} + \frac{\ell}{2}
\]
\[ \mathbf{c}' \mathbf{X} + \frac{c}{2} \]

(1.3.88)

where

\[ \mathbf{c}' = (c_{1r}, \ldots, c_{kr}) \quad \text{and} \quad \bar{X} = \frac{1}{N} \sum_{j=1}^{N} X_j, \]

and

\[ P[|\mathbf{c}' X - \mu| \leq \frac{c}{2} \text{ for all } r] \geq 1 - \alpha \]

(1.3.89)

Problem (5). Finally, Healy considers estimating the mean vector \( \mu \) by a confidence region \( R \) such that the maximum diameter of \( R \) does not exceed \( \frac{c}{2} \). Follow exactly the same sampling scheme as given for problem (4) and estimate \( \mu \) by the set of points \( t \) satisfying

\[ N(\bar{X} - t); S^{-1}(\bar{X} - t) \leq \frac{c^2}{4z}. \]

(1.3.90)

The probability is at least \( 1 - \alpha \) that this set of points includes the true mean \( \mu \).

The nonparametric analogs of Healy's (1956) simultaneous confidence interval procedures were considered by Ghosh and Sen (1971) along with certain robust nonparametric confidence intervals based on the results of Sen (1966) and Sen and Ghosh (1971).

Chatterjee (1962) developed fixed-width confidence intervals, via two-stage sampling, for linear combinations of the elements of a \( p \)-dimensional multivariate normal mean \( \mu \) vector which are independent of the unknown covariance matrix \( \Sigma \). It is based on Chatterjee (1959a).
Let $\mathbf{X} \sim \mathcal{N}_p(\mu, \Sigma)$, and let $A$ be any positive definite matrix.

Take $n_0$ observations on $\mathbf{X}$ and compute $\mathbf{S}$, the usual unbiased estimate of $\Sigma$. Take $N-n_0$ additional observations where

$$N = \max\{n_0 + p^2, \left[\frac{\text{tr}(\mathbf{A}\mathbf{S})}{2}\right] + 1\}$$  \hspace{1cm} (1.3.91)

where $\text{tr}(\mathbf{B})$ is the trace of the matrix $\mathbf{B}$. Let $\mathbf{C} = \begin{pmatrix} \mathbf{C}^1_{p \times N} & \mathbf{C}^2_{p \times N} & \ldots & \mathbf{C}^p_{p \times N} \end{pmatrix}$ be a $p^2 \times N$ dimensional matrix such that

a) the first $n_0$ columns of $\mathbf{C}$ are identical,

b) $\mathbf{C}_\eta = \delta$ where $\eta = (1,1,\ldots,1)'$ and $\delta = (\delta_{11}, \delta_{12}, \ldots, \delta_{1p}, \ldots, \delta_{pp})$,

$c_{ij}$ being Kronecker's delta, and

c) $\mathbf{C}'\mathbf{C} = z\mathbf{A}^{-1} \otimes \mathbf{S}^{-1}$.

Chatterjee (1959a) shows that the matrix $\mathbf{C}$ exists by the choice of $N$ (although he does not give any method for finding $\mathbf{C}$). Let $\mathbf{X}_{p \times N}$ be the final observation matrix and let

$$\tilde{\mu}' = (\text{tr} \mathbf{C}_1 \mathbf{X}', \ldots, \text{tr} \mathbf{C}_p \mathbf{X}')$$

Then $\mu = \frac{1}{z} (\tilde{\mu} - \mu)' \mathbf{A}(\tilde{\mu} - \mu)' \mathbf{A}(\tilde{\mu} - \mu)$ follows the distribution given by (1.3.20) with $\Delta = 0$, and if we choose $u_\alpha$ to be the upper $\alpha$th percent point of the distribution (1.3.20),

$$P\left[\tilde{\mu}' - \sqrt{z u_{\alpha} \xi' \mathbf{A}^{-1} \xi} \leq \tilde{\mu}' \leq \tilde{\mu}' + \sqrt{z u_{\alpha} \xi' \mathbf{A}^{-1} \xi} \quad \forall \xi \in \mathbb{R}^p\right] = 1-\alpha$$  \hspace{1cm} (1.3.92)
If we are interested only in those $\xi$ such that $\xi' A^{-1} \xi = C$ (i.e. contours), then the above gives simultaneous intervals of fixed length for that class of $\xi$. If $A = I$ and $C = 1$ then we are considering the class of all normalized linear functions $\xi' \chi$ such that $\xi' \xi = 1$.

Spjotvoll (1972) presented the following single-sample procedure for simultaneously estimating all linear combinations of population means. Let $X_{ij}$ ($i = 1, \ldots, k$, $j = 1, \ldots, n_i$) be independent normal variates with

$$E(X_{ij}) = \mu_i$$

and

$$\text{Var}(X_{ij}) = \sigma_i^2,$$

$\mu_i$ and $\sigma_i^2$ unknown. Let $\psi = \sum_{i=1}^{k} c_i \mu_i$ be an arbitrary linear combination of the means and let $\hat{\psi} = \sum_{i=1}^{k} c_i \bar{X}_i$ where $\bar{X}_i = 1/n_i \sum_{j=1}^{n_i} X_{ij}$.

Then the variance of $\hat{\psi}$ is $\sigma^2 = \sum_{i=1}^{k} c_i^2 \sigma_i^2/n_i$. An estimate of this variance is $\hat{\sigma^2} = \sum_{i=1}^{k} c_i^2 s_i^2/n_i$, where $s_i^2$ is an estimate of $\sigma_i^2$ such that $(f_i s_i^2)/\sigma_i^2 \sim \chi^2_{f_i}$. Spjotvoll proves, using the Cauchy-Schwarz inequality, that if $A$ is the upper $\alpha$th percent point of the distribution of $Z = \sum_{i=1}^{k} Z(f_i)$ where $\{Z(f_i)\}_{i=1}^{k}$ are independent variates with $Z(f_i) \sim F_{1,f_i}$, then

$$P\left[\frac{\hat{\psi} - A^{1/2}}{\sigma^2} \leq \psi \leq \frac{\hat{\psi} + A^{1/2}}{\sigma^2} \text{ for all } \psi\right] = 1-\alpha \quad \text{(1.3.93)}$$

This procedure then provides simultaneous confidence regions for all linear combinations of the means even if the underlying variances
are unequal but the lengths of the intervals are random.

The exact distribution of $Z$ is unknown and it is suggested that it be approximated by the distribution of $aF_{k,b}$ where $a$ and $b$ are chosen to make the first two moments of $Z$ and $aF_{k,b}$ equal. This approximation becomes exact as the $f_i$ tend to infinity and Morrison (1971) showed the approximation to be excellent for $k = 2$ and $f_1 = f_2$.

Hochberg (1975) employed a two-stage sampling scheme to produce simultaneous confidence intervals for all contrasts and all linear combinations of the means of $k$ independent normal populations with unknown and unequal variances. The procedures depend upon the following distributions and functions. Let $\{t_i\}_{i=1}^k$ be iid Student's-t variates with $n_0 - 1$ degrees of freedom and let $t_0 \equiv 0$ be independent of $\{t_i\}_{i=1}^k$.

Then $Q = \max_{i,j=0,\ldots,k} \{|t_i - t_j|\}$ is distributed as the augmented range of $\{t_i\}_{i=1}^k$, and $Q' = \sum_{i=1}^k (t_i - \bar{t})^2$ (where $\bar{t} = 1/k \sum_{i=1}^k t_i$) is a quadratic form in $\{t_i\}_{i=1}^k$. Let $z = (l_1, \ldots, l_k) \in \mathbb{R}^k$ be an arbitrary vector and let $P = \{i/l_i > 0\}$ and $N = \{i/l_i < 0\}$, and define

$M(z) = M(z_1, \ldots, z_k) = \max\{\sum_{i \in P} l_i, -\sum_{i \in N} l_i\}$.

The two-stage sampling scheme $\mathcal{P}(n_0, 0; s_i^2, z)$ is applied to each population $\pi_i \sim N(\mu_i, \sigma_i^2)$ ($i = 1, 2, \ldots, k$), where $s_i^2$ is the usual unbiased estimate of $\sigma_i^2$ based on the first $n_0$ observations. The overall means $\bar{x}_i$ are computed and it is shown that the probability is $1-\alpha$ that

$$\frac{k}{n_0} \sum_{i=1}^k \bar{x}_i - \sqrt{2} \sum_{i=1}^k c_{k,n_0-1}^{Q} M(b_i, \ldots, b_k) \leq \sum_{i=1}^k l_i \mu_i$$

$$\leq \sum_{i=1}^k l_i \bar{x}_i + \sqrt{2} \sum_{i=1}^k c_{k,n_0-1}^{Q} M(b_i, \ldots, b_k)$$
for all $\ell \in \mathbb{R}^k$, where $b_1 = \ell_i s_i / N_1^{1/2}$, $N_1$ being the final sample-size on $\pi_i$ defined by the sampling procedure and $q^\alpha_{k,n_0-1}$ is the upper $\alpha$th percent point of the distribution of $Q$. It is suggested that in practice $q^\alpha_{k,n_0-1}$ may be replaced by $q^\alpha_{k,n_0-1}$ the upper $\alpha$th percent point of the distribution of the range of $\{t_i\}_{i=1}^k$, the latter providing a satisfactory approximation.

A two-stage analog of Scheffe's S-method is also presented. Consider the same set-up as above.

**Definition (1.3.94).** A contrast in the parameters $\{\mu_i\}_{i=1}^k$ is a linear combination $\sum_{i=1}^k c_i \mu_i$ such that $\sum_{i=1}^k c_i = 0$.

It is shown that the probability is at least $1-\alpha$ that all contrasts satisfy

$$\sum_{i=1}^k c_i \tilde{x}_i - \sqrt{z} q^\alpha_{k,n_0-1} (\sum_{i=1}^k c_i^2)^{1/2} \leq \sum_{i=1}^k c_i \mu_i$$

$$\leq \sum_{i=1}^k c_i \tilde{x}_i + \sqrt{z} q^\alpha_{k,n_0-1} (\sum_{i=1}^k c_i^2)^{1/2}$$

where $q^\alpha_{k,n_0-1}$ is the upper $\alpha$th percent point of the distribution of $Q'$.

This may be extended to all linear combinations if $q^\alpha_{k,n_0-1}$ is replaced by $q^\alpha_{k,n_0-1}$ the upper $\alpha$th percent point of the distribution of $\tilde{Q} = \sum_{i=1}^k t_i^2$.

The distribution of $Q'$ has not been tabulated, and Hochberg suggests approximating it by that of $F_{\ell,m}$ where $\ell$ and $m$ are
implicitly defined by equating the first two moments of \( Q' \) and \( IF_{l,m} \).

These moments are

\[
E(Q') = \frac{(k-1)n_0^{-1}}{n_0 - 2}, \quad (1.3.96)
\]

\[
E(Q'^2) = \frac{(n_0^{-1})^2(k-1)}{(n_0 - 3)n_0 - 4 + \frac{k-2k+5}{n_0 - 3}} \quad (1.3.97)
\]

\[
E(IF_{l,m}) = \frac{n}{m-2}, \quad (1.3.98)
\]

and

\[
E(IF_{l,m}^2) = \frac{m^2 l(l+2)}{(m-2)(m-4)} \quad (1.3.99)
\]

A numerical example is considered and graphs are given for finding the upper 20% critical points for the ranges of \( k \) i.i.d. Student's-t variates \( k = 2(1)10, n_0 = 6, 9, 11, 16, 21, 31, 41, 61 \). A table is given with approximate values of \( q_{\alpha k,n_0-1}^\alpha \) for \( \alpha = .05, .10, k = 3(1)10 \), and the range of \( n_0 \) as above.

Dudewicz, Ramberg and Chen (1975) solved the problem of simultaneously comparing \( k-1 \) populations with a control when each population is normally distributed with unknown means and variances. That is, let \( \pi_1 \sim \mathcal{N}(\mu_1,\sigma_1^2) \) be the control population and \( \pi_i \sim \mathcal{N}(\mu_i,\sigma_i^2) (i = 2,\ldots,k) \) the competitors to the control. Then one-sided confidence intervals are obtained on \( (\mu_i - \mu_1) (i = 2,\ldots,k) \) with an overall confidence coefficient, \( 1-\alpha \). Exact tables are given to implement the procedure.

The two-stage procedure \( P(n_0,1;\sigma_i^2,d) \) is used on population \( \pi_i \), where \( \sigma_i^2 \) is the usual unbiased estimate of \( \sigma_i^2 \), based on the first \( n_0 \) observations and \( d \) is the solution of the equation
where \( F_{n_0}(\cdot) \) and \( f_{n_0}(\cdot) \) are respectively the distribution function and density function of the Student's-t distribution with \( n_0-1 \) degrees of freedom and \( a \) is the smallest difference the experimenter wishes to detect.

Let \( \tilde{X}_i(n_0) = \sum_{j=1}^{n_0} x_{ij}/n_0 \) and \( Y(N_i-n_0) = 1/(N_i-n_0) \sum_{j=n_0+1}^{n_1} x_{ij}/(N_i-n_0) \), where \( N_i \) is the final sample size for population \( \pi_i \). Set

\[
\tilde{X}_i = c_i \tilde{X}_i(n_0) + (1-c_i) \tilde{Y}(N_i-n_0) \quad (i = 1, \ldots, k)
\]

where

\[
c_i = \frac{n_0}{N_i} \left( 1 + \sqrt{1 - \frac{N_i}{n_0} \left( 1 - \frac{(N_i-n_0)}{(s_i^2/d_i^2)} \right)} \right) . \tag{1.3.101}
\]

Then the probability is \( 1-\alpha \) that

\[
\mu_i - \mu_1 \leq (\tilde{X}_i - \tilde{X}_1) + a \quad (2 \leq i \leq k) , \tag{1.3.102}
\]

which gives simultaneous upper intervals; and similarly the probability is \( 1-\alpha \) that

\[
(\tilde{X}_1 - \tilde{X}_i) - a \leq \mu_i - \mu_1 \quad (2 \leq i \leq k) , \tag{1.3.103}
\]

which gives the corresponding lower intervals.

The same problem of multiple-comparison with a control was considered earlier by Dudewicz and Ramberg (1972) in which the two-stage
procedure and approximate tables (derived by linear interpolation) were presented.

Andrews and Arnold (1974) worked on the problem of finding fixed-width confidence intervals for the mean of a normal population when it is assumed to lie in an interval \((a, b)\). Single-stage, two-stage, and sequential procedures are developed through a Bayesian framework which assumes that the unknown mean \(\mu\) has a uniform prior on \((a, b)\). The cases of \(\sigma^2\) known and \(\sigma^2\) unknown are both considered. These procedures become the classical procedures for one, two and sequential-stage estimation as \(a \to -\infty\) and \(b \to +\infty\).

We next review the literature which deals specifically with analysis of variance when the cell variances are unequal. Until now only single-sample procedures have appeared in the literature. Many of these procedures rely on approximations to the distributions of the test statistics, and all depend upon the unknown variances. In many cases the authors depend upon "robustness" properties of the procedures to claim that the heteroscedasticity is "not significant".

The reviews of this section provide incentive to look beyond one-stage procedures to develop tests with known distributions which are completely independent of the variances. Such procedures will then be completely robust against inequality of variances!

Let \(X_1, \ldots, X_k\) be random variables which are distributed normally and independently with means \(\mu_1, \ldots, \mu_k\) and variances \(\sigma^2_1, \ldots, \sigma^2_k\) respectively. It is desired to test the hypothesis

\[
H_0 : \mu_1 = \cdots = \mu_k
\]  

(1.3.104)
Let \( s_1^2, \ldots, s_k^2 \) be estimates of \( \sigma_1^2, \ldots, \sigma_k^2 \) based on \( n_1, \ldots, n_k \) degrees of freedom respectively. If the variances were known the test of (1.3.104) would be based on the statistic

\[
\sum_{i=1}^{k} w'_i (X_i - \bar{X})^2
\]

(1.3.105)

where \( w'_i = 1/\sigma_i^2 \), \( w = \sum_{i=1}^{k} w'_i \) and \( \bar{X} = \sum_{i=1}^{k} w'_i X_i / w' \); under \( H_0 \) (1.3.105) is distributed as \( \chi^2_{k-1} \).

If the variances are unknown replace \( w'_i \) by \( w_i = 1/s_i^2 \) and \( w \) by \( w = \sum_{i=1}^{k} w_i \), (note that if the \( X_i \)'s are sample means based on \( n_i \) observations then \( w_i = n_i / s_i^2 \)) if the \( r_i \)'s are "large" assume that

\[
\sum_{i=1}^{k} w_i (X_i - \bar{X})^2
\]

(1.3.106)

is distributed approximately as \( \chi^2_{k-1} \).

If the \( r_i \)'s are not "large", James (1951) suggests seeking a function \( h(s_1^2, \ldots, s_k^2) \) such that

\[
P[ \sum_{i=1}^{k} w_i (X_i - \bar{X})^2 \leq h(s_1^2, \ldots, s_k^2) ] = \alpha
\]

(1.3.107)

Following Welch (1947), a series solution for \( h \) in (1.3.107) is sought and it is shown that (to the order of \( 1/r_i \)) \( h \) may be expressed as

\[
h(w_1, \ldots, w_k) = \chi^2_{k-1, (1-\alpha)} \left[ 1 + \frac{3\chi_{k-1, (1-\alpha)} + (k+1)}{2(k^2 - 1)} \cdot \sum_{i=1}^{k} \frac{1}{r_i} \left( 1 - \frac{w_i}{w} \right)^2 \right]
\]

(1.3.108).
The test of (1.3.104) is based on the statistic (1.3.106) and the probability statement given by (1.3.107), and has the property that the level of the test is approximately $\alpha$ and hence is robust against inequality of variances.

*Welch (1951)* considered the same problem as James (1951) but provided a different function $h(s_1^2, \ldots, s_k^2)$, namely,

$$h'(s_1^2, \ldots, s_k^2) = F_{1-\alpha}^{K-1, f} \left[ 1 + \frac{2(k-2)}{k^2-1} \sum_{i=1}^{k} \frac{1}{r_i} \left( 1 - \frac{w_i}{w} \right)^2 \right] \quad (1.3.109)$$

where $f$ is implicitly defined as

$$\frac{1}{f} = \left( \frac{3}{(k^2-1)} \right) \sum_{i=1}^{k} \frac{1}{r_i} \left( 1 - \frac{w_i}{w} \right)^2. \quad (1.3.110)$$

Tests of hypotheses for the general linear model under the assumption of homoscedastic variances are well known. *James (1954)* assumes a general linear model with heteroscedastic variances and considers tests of linear hypotheses.

Let $X_1, X_2, \ldots, X_n$ be $n$ random variables whose mean values are linear functions of $\theta_1, \ldots, \theta_k$ and which are distributed independently and normally with variances $\sigma_1^2, \ldots, \sigma_n^2$. Hence

$$E(X) = B\varnothing,$$

where $X$ and $\varnothing$ are the vectors $(X_1, \ldots, X_n)$ and $(\theta_1, \ldots, \theta_k)$ and $B$ is an $n \times k$ matrix of known constants $b_{ij}$. We write $\varnothing = (\varnothing_1, \varnothing_2)$ where $\varnothing_1$ is an $r$ dimensional vector $(\theta_1, \ldots, \theta_r)$ and $\varnothing_2$ is a
k-r dimensional vector \( (\theta_{r+1}, \ldots, \theta_k) \), and desire to test
\[
H_0 : \theta \sim_1 = 0 .
\] (1.3.111)

If the \( \sigma_i^2 \)'s were known the likelihood ratio test would be based on the statistic
\[
T(\sigma_1, \ldots, \sigma_n) = \frac{\hat{\Theta}^t V^{-1} \hat{\Theta}}{\sim_1} \sim_1
\] (1.3.112)

where \( \hat{\Theta} = c^{-1}d \), \( c = B'D^{-1}B \), \( D = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2) \) and \( V \) is the variance covariance matrix of \( \hat{\Theta} \). Under \( H_0 \), \( T(\sigma_1, \ldots, \sigma_n) \) is distributed as \( \chi^2_r \). Since the \( \sigma_i^2 \)'s are not known but their estimates \( \hat{s}_i^2 \) are available and noting that (1.3.112) is a function of \( \sigma_1^2, \ldots, \sigma_k^2 \), James bases a test of \( H_0 \) on the statistic (1.3.112) with the \( \sigma_i^2 \)'s replaced by the \( s_i^2 \)'s and uses the approximation
\[
P[T(s_1, \ldots, s_n) \leq \chi^2_{r, \alpha}] = \alpha .
\]

He then seeks a function \( h(s_1^2, \ldots, s_k^2) \) (as in James (1951)) such that
\[
P[T(s_1, \ldots, s_n) \leq h(s_1, \ldots, s_n)] = \alpha .
\]

As the sample size approaches infinity \( h(s_1, \ldots, s_n) \) will approach \( \chi^2_{r, \alpha} \). The function \( h \) is obtained as an infinite series.

Examples and applications are considered for several linear hypotheses, including the one-way layout and the two-way layout, and extensions to multivariate linear hypotheses are presented.
Several procedures have been proposed to test equality of means in a one-way layout with unequal variances. Brown and Forsythe (1974a) study the small-sample behavior of four of these in terms of level of significance and power based on a Monte Carlo sampling experiment.

Let $X_{ij} \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, k; j = 1, \ldots, n_i$. The four procedures studied utilized:

1) The usual F-statistic

$$F = \frac{\sum_{i=1}^{k} \frac{n_i (\bar{X}_i - \bar{X})^2}{(k-1)}}{\sum_{i=1}^{k} (n_i-1) s_i^2 / (\sum_{i=1}^{k} n_i - k)}$$

(1.3.113)

where

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij},$$

$$\bar{X} = \sum_{i=1}^{k} \frac{n_i}{\sum_{i=1}^{k} n_i} \bar{X}_i,$$

$$s_i^2 = \sum_{j=1}^{n_i} \frac{(X_{ij} - \bar{X}_i)^2}{n_i - 1};$$

2) Welch's (1951) statistic

$$W = \frac{\sum_{i=1}^{k} w_i (\bar{X}_i - \bar{X})^2/(k-1)}{[1 + \frac{2(k-2)}{(k-1)} \sum_{i=1}^{k} (1 - w_i/u)^2/(n_i - 1)]}$$

(1.3.114)

where

$$w_i = \frac{n_i}{s_i^2},$$

$$u = \sum_{i=1}^{k} w_i,$$
\[ \bar{X} = \sum_{i=1}^{k} w_i \bar{X}_i / u \]

(Under the null hypothesis of equality of means \( W \) is approximately distributed as \( F_{k-1,f} \) where \( f \) is implicitly defined by

\[ 1/f = (3/(k^2-1)) \sum_{i=1}^{k} (1 - w_i/u)^2/(n_i-1) \];

3) James' (1951) statistic

\[ J = \sum_{i=1}^{k} w_i (\bar{X}_i - \bar{X})^2 / k-1 \]

(1.3.115)

If all of the means are equal then

\[ F \left[ J > \chi^2_{k-1,1-2} \left[ 1 + \frac{3\chi^2+k+1}{k} \sum_{i=1}^{k} (1 - \frac{w_i}{n})^2/(n_i-1) \right] \right] = \alpha ; \]

and

4) Brown and Forsythe's statistic

\[ F^* = \frac{\sum_{i=1}^{k} n_i (\bar{X}_i - \bar{X})^2}{\sum_{i=1}^{k} (1-n_i/\bar{X}_i^k \sum_{i=1}^{k} n_i s_i^2)} \]

(under the null hypothesis of equality of means \( F^* \) is approximately distributed as \( F_{k-1,f} \) where \( f \) is implicitly defined by

\[ 1/f = \sum_{i=1}^{k} c_i^2/(n_i-1) \]

)
with

\[ c_i = \frac{(1-n_i/\sum_{i=1}^{k} n_i) s_i^2 / (\sum_{i=1}^{k} (1-n_i/\sum_{i=1}^{k} n_i) s_i^2))}{\sum_{i=1}^{k} (1-n_i/\sum_{i=1}^{k} n_i) s_i^2))} \]

A Monte Carlo sampling experiment was performed under varying circumstances (i.e. different configurations of means, variances, and sample sizes). For each configuration, the Monte Carlo experiment was repeated 10,000 times, and each statistic was computed and compared to its nominal 90, 95 and 99th percentiles. The corresponding levels and power were then estimated.

Tables of the results are presented and indicate the sensitivity of \( F \) to unequal variances. For example if an experimenter thinks he is running a nominal 10% test with 4 populations and if 4, 8, 10, 12 are the respective sample sizes and 1, 2, 2, 3 the respective standard deviations, the true level is approximately .059. The asymptotic approximations of \( F^* \) and \( W \) appear valid when each group has at least 10 observations. It is suggested that \( F^* \) or \( W \) be used depending upon whether extreme means are thought to have extreme variances (use \( W \)) or not (use \( F^* \)). If some sample variances appear unusually low \( F^* \) is suggested.

Brown and Forsythe (1974a) studied the robustness properties of the statistic \( F^* \) defined there. Brown and Forsythe (1974b) show \( F^* \) may be derived by combining individual orthonormal contrasts which are defined below.

Set \( X_{ij} = \mu_i + e_{ij}, i = 1, \ldots, k; j = 1, \ldots, n_i, \) \( \{e_{ij}\} \) independent with \( e_{ij} \sim N(0, \sigma_i^2) \). Set
A set of orthonormal contrasts is

\[ C = \{ \sum_{j=1}^{k} d_{ij} \mu_j, \ i = 1, \ldots, L \} \]

where

\[ \sum_{j=1}^{k} d_{ij} = 0, \quad (i = 1, \ldots, L) \]

\[ \sum_{j=1}^{k} d_{ij}^2/n_j = 1, \quad (i = 1, \ldots, L) \]

\[ \sum_{j=1}^{k} d_{ij} d_{mj}/n_j = 0, \quad (i = 1, \ldots, L; \ i \neq m) \]

A statistic which jointly tests that each of these \( L \) contrasts is zero is

\[ F_L^* = \frac{\sum_{i=1}^{L} \sum_{j=1}^{k} d_{ij}^2 \bar{x}_{ij}^2}{\sum_{i=1}^{L} \sum_{j=1}^{k} d_{ij}^2 s_j^2/n_j} \]

which is (under the hypothesis that each of the \( L \) contrasts is zero) approximately distributed as \( F_{L,f} \) where \( f \) is implicitly defined by

\[ \frac{1}{f} = \frac{\sum_{j=1}^{k} c_{ij}^2}{(n_j - 1)}, \quad c_j = \frac{\sum_{i=1}^{L} d_{ij}^2 s_j^2/n_j}{\sum_{j=1}^{L} \sum_{i=1}^{k} d_{ij}^2 s_j^2/n_j} \]
It is shown that when $L = k-1$, $F_L^*$ is equivalent to $F^*$. This further justifies $F^*$ as a single-sample solution to the one-way ANOVA problem (since in that case the hypothesis that each of the $L$ contrasts is zero is equivalent to the hypothesis that all $k$ means are equal).

Similar consideration is given to the two-way layout and to multiple-comparisons.

Mazuy and Connor (1965) investigated the problem of testing contrasts for row and column means in a two-way layout when the cell variances are unknown and unequal. The method discussed may be considered a generalization of Scheffe (1943) in that independent estimates of the contrasts to be tested are used in calculating a $t$-statistic used in the test.

The model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + e_{ijk} \quad (1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq n)$$

with

$$e_{ijk} \sim N(0, \sigma^2_{\alpha_i} + \sigma^2_{\beta_j}) \quad \sum_{i=1}^{I} \alpha_i = 0 \quad \text{and} \quad \sum_{j=1}^{J} \beta_j = 0.$$

(The model with interactions is also considered.)

A general theory is presented, letting

$$C_s(A) = \sum_{i=1}^{I} a_{si}Y_{i..} \quad \sum_{i=1}^{I} a_{si} = 0$$

($Y_{i..}$ denotes the average for row $i$)

$$C_t(B) = \sum_{j=1}^{J} b_{tj}Y_{..j} \quad \sum_{j=1}^{J} b_{tj} = 0$$
\(Y_{ij}\). denotes the average of the \(j\)th column).

\[
C_{st}(AB) = \sum_{i=1}^{I} \sum_{j=1}^{J} a_i b^T j Y_{ij}, \quad \sum_{i=1}^{I} a_i = 0, \quad \sum_{j=1}^{J} b_j = 0
\]

\(Y_{ij}\). denotes the average for cell \(ij\), and

\[
C_r(E) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{n} c_{rijk} Y_{ijk}, \quad \sum_{i} \sum_{j} \sum_{k} c_{rijk} = 0
\]

denote contrasts among rows, columns, interactions and error.

The hypothesis of interest is

\[
H_0 : E(C_s(A)) = 0 \quad (1.3.117)
\]

and it is desired to base the test of (1.3.117) on a Student's-t statistic. \(C_s(A)/C(E)\) will be such a statistic provided:

1) \(\text{cov}(C_s(A), C(E)) = 0\)

2) \(\text{V}(C_s(A)) = K \text{V}(C(E))\),

where \(K > 0\) is independent of \(\sigma^2_{\alpha_i}\) and \(\sigma^2_{\beta_j}\).

Similar statistics can be constructed for tests of column contrasts. These restrictions generate a class of linear functions of the form \(\sum_{i,j,k} c_{rijk} Y_{ijk}\) which may be used as the denominator of the t-statistics. Properties of this class (i.e. degrees of freedom, basis, etc.) are discussed and it is shown that if \(C(E)\) is in the class it must be a linear combination of interaction contrasts and error contrasts. A numerical example illustrates the application of the procedure.
Chakravarti (1965) extended the solutions of Scheffé (1943) and Bartlett (as referred to by Neyman (1941)) to the Behrens-Fisher problem to more than two populations. A class of tests for the hypothesis of equality of means is given and we will describe in detail that test which is optimal in the sense that the estimates of the means have minimum variance.

Let $\pi_1$, $\pi_2$, ..., $\pi_k$ be $k \geq 2$ populations such that $\pi_i \sim N(\mu_i, \sigma_i^2)$. It is desired to test the null hypothesis

$$H_0 : \mu_1 = \cdots = \mu_k$$  \hspace{1cm} (1.3.118)

Let $(x_{i1}, ..., x_{in_i})$ be a random sample of size $n_i$ from $\pi_i$ and assume without loss of generality $n_1 \leq n_2 < \cdots < n_k$. Compute the Scheffé variables (see 1.3.39)

$$U_{j\alpha} = X_{1\alpha} - \sum_{r=1}^{n_j} b_{j\alpha r} X_{r\alpha}$$  \hspace{1cm} (1.3.119)

where

$$\sum_{r=1}^{n_j} b_{j\alpha r} = 1 \quad \text{and} \quad \sum_{r=1}^{n_j} b_{j\alpha r} b_{j\alpha' r} = b_{j}^2 \delta_{\alpha\alpha'}$$  \hspace{1cm} (1.3.120)

and $b_{j}^2$ is a constant with $\delta_{ij}$ being Kronecker's delta. In this case

$$E(U_{j\alpha}) = \mu_i - \mu_j = \delta_{ij}$$

and

$$\text{Var}(U_{j\alpha}) = \sigma_i^2 + b_{j}^2 \sigma_j^2,$$
and using Scheffe's optimal choice for the $b_{j\alpha}$ (see equation (1.3.41)) we have

$$\text{Var}(U_{j\alpha}) = \sigma_i^2 + \frac{n_i}{n_j} \sigma_j^2. \quad (1.3.122)$$

It follows that $U_{\sim\alpha} = (U_{2\alpha}, \ldots, U_{k\alpha})$ follows a multivariate normal distribution $N_{k-1}(\delta, \Sigma)$ where

$$\delta = (\mu_{1\alpha} - \mu_{2\alpha}, \mu_{1\alpha} - \mu_{3\alpha}, \ldots, \mu_{1\alpha} - \mu_{k\alpha}), \quad (1.3.123)$$

and

$$\sigma_{ij} = \begin{cases} 
\sigma_i^2 + \frac{n_i}{n_j} \sigma_j^2, & i = j \\
\sigma_i^2, & i \neq j.
\end{cases} \quad (1.3.124)$$

We then take $U_{\sim 1}, \ldots, U_{\sim n_1}$ as a random sample from this distribution and note that hypotheses (1.3.118) is equivalent to $H_0: \delta = \Omega$ for the generated multivariate normal distribution. We may test $H_0: \delta = \Omega$, and hence (1.3.118), using the statistic

$$\frac{T^2}{n_1 - 1} = n_1 \bar{U}^T (S^{-1}) \bar{U}. \quad (1.3.125)$$

where $\bar{U} = 1/n_1 \sum_{\alpha=1}^{n_1} U_{\alpha}$ and $S = (S_{ij})$ where

$$S_{ij} = \sum_{\alpha=1}^{n_1} (U_{i\alpha} - \bar{U}_i) (U_{j\alpha} - \bar{U}_j) \quad (\bar{U}_i = 1/n_1 \sum_{\alpha=1}^{n_1} U_{i\alpha}), \quad (1.3.126)$$

which is distributed as Hotelling's $T^2$ with $n_1 - 1$ degrees of freedom for a $(k-1)$-dimensional distribution. The distribution of $T^2/(n_1 - 1)$
under alternative hypotheses is $F_{k-1,n_1-k+1}(\Delta)$, where $\Delta = n_1 \delta^2 \mathbf{I}^{-\frac{1}{2}}$.
(Note the power of this test is not independent of the unknown variances.)

It is clear that we may choose any of the $k$ populations to play the role of $\pi_1$ above. If $\pi_1$ is used then $E(U_{j\alpha}) = \mu_i - \mu_j$ and $\text{var}(U_{j\alpha}) = \sigma_i^2 + b_j^2 \sigma_j^2$. Thus the above procedure is one representative from a class of tests for (1.3.118). However, when populations other than $\pi_1$ are used, observations must be discarded since $n_j \geq n_1$ and a mixture of Bartlett's and Scheffe's variables must be used. It is shown that in terms of power it is best to use the procedure with $\pi_1$ (i.e. the population with the smallest sample-size).

A comparison is made between Bartlett's test and Scheffe's test for the Behrens-Fisher problem and Scheffe's procedure is shown to be superior when the sample sizes are unequal. (In the case of equal sample sizes the procedures are equivalent.)

Problems with unequal variances arise naturally when two independent experiments are performed on each of $k$ treatments, at different locations. Kulkarni (1973) considers two hypothesis testing problems in this heteroscedastic situation. The model for this problem is given by

$$X_{ij} = \mu_i + e_{ij} \quad i = 1, \ldots, k, \ j = 1, \ldots, n_1$$

which are the observed random variables from the first location and

$$Y = \mu_i^* + e_i'_{ij} \quad i = 1, \ldots, k, \ j = 1, \ldots, n_2$$

which are the corresponding observations from the second location. $\{e_{ij}\}$ are i.i.d. with $e_{ij} \sim N(0, \sigma_1^2)$ and the $\{e_i'_{ij}\}$ are i.i.d. with $e_i'_{ij} \sim N(0, \sigma_2^2)$. 
Expressing $\mu_i$ and $\mu_i^*$ as

$$\mu_i = \mu + \lambda_i + I_{il}$$

and

$$\mu_i^* = \mu_i + \lambda_i + I_{i2}$$

where $\lambda_i$ is called the effect of the $i$th treatment and $I_{ij}$ is called the interaction between the $i$th treatment and $j$th location the hypotheses of interest are

1) $H_0 : \lambda_i = 0$ $i = 1,...,k$ (1.3.127)

2) $H_1 : I_{il} = I_{i2}$ $i = 1,...,k$ (1.3.128)

Tests known as "bilateral" tests have been proposed in the literature to test (1.3.127) and (1.3.128). Kulkarni studies the small-sample properties of these tests in terms of their level and power as functions of $R = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$. Some power properties are obtained and two tests equivalent to bilateral tests for large samples and which are better than bilateral tests for small samples are presented.

Demskey (1975) is concerned with the problem of evaluating three or more means under conditions of heteroscedasticity. His claim that "the methods given to date evaluate only means for two group experiments; no method is given to evaluate 3 or more groups, batches or lots..." seems incorrect in light of the papers by James, Welch, Chakravarti and others reviewed previously.

Four methods, Welch-Aspin, Behrens-Fisher, Cochran and INTERPRO (discussed in the paper), are compared for $k = 2$. A heuristic method
is suggested based on the INTERPRO method for 3 or more groups.

Box and Hill (1974) were interested in a linear model with unequal variances and presented a method for obtaining approximate weights in a weighted least squares analysis where the variance of the dependent variable is a function of the mean. Specifically let \( \mathbf{y} = (Y_1', \ldots, Y_n')' \) be an \( n \times 1 \) vector of independent observations. Set \( \mathbf{\theta} = (\theta_1', \ldots, \theta_p')' \), and

\[
Y_i = \eta_i + e_i \quad i = 1, \ldots, n
\]

where the \( \eta_i \) are linear functions of \( \mathbf{\theta} \) and \( e_i \sim N(0, c_i \sigma^2) \), the \( c_i \) being unknown. Suppose there exists some unknown power transformation \( Y_i^{(\varphi)} \) defined by

\[
Y_i^{(\varphi)} = \begin{cases} 
Y_i^{\varphi-1} / \varphi, & \varphi \neq 0 \\
\log Y_i, & \varphi = 0
\end{cases}
\]

which has a constant variance.

An approximate variance expression is developed, a procedure for estimating \( \varphi \) given, and the approximate weighted least-squares analysis performed on the transformed data (which now has approximately equal variances). The methods also apply to nonlinear models. Finally a numerical example is presented.

The generalization of the analysis of variance problems with unknown variances to the multivariate analysis of variance with unknown covariance matrices has been considered by several authors.
Bennett (1951) extended Scheffé's (1943) procedures to test the equality of mean vectors of two p-dimensional multivariate normal distributions when the covariance matrices are unknown. Anderson (1963) generalizes these results to more than two multivariate populations when equality of mean vectors is to be tested.

Let \( X_{ij} \sim N_p(\mu_{i}, \Sigma_i) \) \( (i = 1, \ldots, k, j = 1, \ldots, n_i) \). As an example Anderson sets \( k = 3, n_1 = n_2 = n \) and defines

\[
Y_{ij} = a_1 X_{1j} + a_2 X_{2j} + a_3 X_{3j}
\]

\[
Z_{ij} = b_1 X_{1j} + b_2 X_{2j} + b_3 X_{3j}
\]

where \( \Sigma_{i=1}^{3} a_i = 0, \Sigma_{j=1}^{3} b_j = 0 \), and \((a_1, a_2, a_3) \) and \((b_1, b_2, b_3) \) are linearly independent. The hypothesis of equality of mean vectors is equivalent to

\[
E(Y_{ij}) = 0, \quad E(Z_{ij}) = 0,
\]

the covariance matrix of \((Y_{ij}, Z_{ij})\) is

\[
\begin{pmatrix}
a_1^2 \Sigma_{11} + a_2^2 \Sigma_{22} + a_3^2 \Sigma_{33} & a_1 b_1 \Sigma_{11} + a_2 b_2 \Sigma_{22} + a_3 b_3 \Sigma_{33} \\
a_1 b_1 \Sigma_{11} + a_2 b_2 \Sigma_{22} + a_3 b_3 \Sigma_{33} & b_1^2 \Sigma_{11} + b_2^2 \Sigma_{22} + b_3^2 \Sigma_{33}
\end{pmatrix}
\]

and \((1,3,140)\) may be tested using
\[ T^2 = n(\bar{Y}', \bar{Z}')S^{-1}(\bar{Y}) \]

where \( \bar{Y} = 1/n \sum_{j=1}^{n} Y'_j, \bar{Z} = 1/n \sum_{j=1}^{n} Z_j \) and

\[ S = \frac{1}{n-1} \sum_{j=1}^{n} \left( \begin{array}{c} Y_j - \bar{Y} \\ Z_j - \bar{Z} \end{array} \right) \left( \begin{array}{c} Y_j - \bar{Y} \\ Z_j - \bar{Z} \end{array} \right) \]

Under \( H_0 : \mu_1 = \mu_2 = \mu_3 \)

\[ \frac{(n-2p)T^2}{(n-1)2p} \sim F_{2p, n-2p} \]

Bhargava (1971) worked on the same problem as Anderson (1963), and in fact is a different generalization of Bennett (1951). As in Anderson (1963) let

\[ X_{ij} \sim N_p(\mu_i, \Sigma_i) \quad (i = 0, 1, \ldots, k, \ j = 1, \ldots, n_i) \]

and \( n = n_1 = \cdots = n_k = n_0/k. \) We want to test

\[ H_0 : \mu_0 = \cdots = \mu_k. \] (1.3.130)

Construct the new variables

\[ Y_{1j} = X_{1j} - X_{0j}, \ Y_{2j} = X_{2j} - X_{0j} + n', \cdots \]

\[ Y_{ij} = X_{ij} - X_{0j} + (j-1)n', \ Y_{kj} = X_{kj} - X_{0j} + (k-1)n \]

Now \( Y_{11}, Y_{12}, \ldots, Y_{1n} \) are independent observations from \( N_p(\mu_1 - \mu_0, \Sigma_1 + \Sigma_0) \) and (1.3.130) becomes
\[ H_0 : \mu_i - \mu_0 = 0 , \quad i = 1, \ldots, k \]

Setting \( H_0^i = \mu_i - \mu_0 = 0 \), (1.3.144) can also be defined by

\[ H_0 = H_0^1 \wedge H_0^2 \wedge \cdots \wedge H_0^k \]

(i.e. the simultaneous holding of the \( H_0^i \) hypotheses). Let \( \lambda_i \) be the likelihood-ratio tests statistic for \( H_0^i \). Then (if \( T_1^2 \) is the Hotelling's \( T^2 \) statistic for the one-sample problem \( H_0^i : \mu_i - \mu_0 = 0 \)) we have

\[ \lambda_i = \left( \frac{1}{1 + \frac{T_1^2}{(n-1)}} \right) \]

If \( \lambda \) is the likelihood-ratio statistic for (1.3.130), then

\[ \lambda = \prod_{i=1}^{k} \lambda_i \]

and the test of (1.3.130) may be based on \( -2 \ln \lambda \), since (see pages 203-210 of Anderson (1958))

\[ P[-2 \ln \lambda \leq M] \approx 1 - \omega_2 \frac{1}{p} \left[ \frac{\chi^2}{p(k-1)} < \rho M \right] + \omega_2 \frac{1}{p} \left[ \frac{\chi^2}{p(k-1)+4} \leq \rho M \right] \]

where

\[ \rho = 1 - \frac{1}{2} \left( \frac{p+2}{n} \right) , \]

\[ \omega_2 = \frac{(k-1)}{48\rho^2 n^2} p(p^2-4) . \]

Note that when \( k = 1 \) this test reduces to that of Bennett (1951), as does that of Anderson (1963).
We noted previously that two-stage procedures have been applied to other areas of statistics. One very successful application has been to the area of ranking and selection which is closely allied with analysis of variance.

Let \( \pi_1, \ldots, \pi_k \) be \( k \geq 2 \) populations such that \( \pi_i \sim N(\mu_i, \sigma_i^2) \), and let \( \mu[1] \leq \mu[2] \leq \cdots \leq \mu[k] \) be the ranked means. The goal is to select the population with the largest mean \( \mu[k] \), and we seek a procedure which yields a probability of correct selection, \( P[CS] \), which is at least \( P^* \) whenever \( \mu[k] - \mu[k-1] \geq \delta^* \), where \( 0 < \delta^* \) and \( 1/k < P^* < 1 \).

Dudewicz (1971) has shown that no single-sample procedure for a selection problem with unknown variances can satisfy this requirement and Dudewicz and Dalal (1975) developed the following two-stage procedure which meets the above requirements.

The two-stage sampling rule \( R(n_0, l; s^2, \delta^*/h) \) is used for population \( \pi_i \), where \( s^2 \) is the usual unbiased estimate of \( \sigma_i^2 \) based on the first \( n_0 \) observations and \( h \) is the unique solution of the integral equation

\[
\int_{-\infty}^{\infty} F_{n_0}^{-1}(x+h) f_{n_0}^{-1}(x) \, dx = P^* \quad (1.3.131)
\]

where \( F_{n_0}^{-1}(\cdot) \) and \( f_{n_0}^{-1}(\cdot) \) are respectively the CDF and density function of a \( t_{n_0-1} \) random variate. Next chose constants \( \{a_{ij}\} \) such that

\[
(a_{11} = \cdots = a_{in_0}) \quad (1.3.132a)
\]
where \( N_1 \) is the final sample size on population \( \pi_1 \). Compute the generalized sample means

\[
\tilde{x}_i = \frac{1}{N_1} \sum_{j=1}^{N_1} a_{ij} x_{ij} \quad (i = 1, \ldots, k),
\]

where \( (x_{i1}, \ldots, x_{iN_1}) \) is the total sample from \( \pi_1 \). Now select the population which yielded \( \tilde{x}_{[k]} \) where \( \tilde{x}_{[1]} \leq \cdots \leq \tilde{x}_{[k]} \) are the order values of \( \tilde{x}_1, \ldots, \tilde{x}_k \). The above procedure is called \( P_E \) and it is shown that the \( P[CS] \) using \( P_E \), denoted by \( P[CS|P_E] \), is independent of \( c_1^2, \ldots, c_k^2 \) and satisfies

\[
\inf\{P[CS|P_E]: \mu_{[k]} - \mu_{[k-1]} \geq \delta^*\} = \frac{P^*}{P^*}.
\]

A second procedure \( P_R \) is considered which is similar to \( P_E \) except that \( \tilde{x}_i \)'s, the overall sample means, are used rather than \( \tilde{x}_i \)'s, and the sampling procedure \( P(n_0, 0; s_i^2, \delta^*/h) \) is used. Comparisons between the two procedures are studied explicitly for \( k = 2 \), and it is shown that

\[
P[CS|P_R] \geq P[CS|P_E] ;
\]  

in fact,

\[
P[CS|P_R, s_1^2, \ldots, s_k^2] \geq P[CS|P_E, s_1^2, \ldots, s_k^2]
\]

uniformly in \( s_1^2, \ldots, s_k^2 \).
For the case $k > 2$ it is shown that (1.3.133) does not hold uniformly in $\varepsilon_1^2, \ldots, \varepsilon_k^2$ and the question of the superiority of $\mathcal{P}_R$ in general is left open (and was considered by Rinott (1975)).

Applications of the above to simulation are considered. Procedures for the problem of subset selection (or, in fact, any ranking and selection goal) are given. Extensive tables, needed to carry out the selection procedure $\mathcal{P}_E$, are provided.

Rinott (1975) proposed to modify Dudewicz and Dalal's procedure $\mathcal{P}_R$ by determining $h$ by the equation

$$
\int_0^\infty \int_0^\infty \frac{h}{(n_0-1)(1/x+i/y)^{1/2}} \phi f(x)dx f(y)dy = \sigma_0^*(1.3.134)
$$

where $f(x)$ is the density of a $\chi^2_{n_0-1}$ random variable. He claims that then $\inf \{ P[CS|\mathcal{P}_R(h):\mu_k - \mu_{k-1} \geq \delta^*] \geq \sigma^*_0 \}$ for all values of $\varepsilon_1^2, \ldots, \varepsilon_k^2$.

Properties of procedure $\mathcal{P}_R$ are considered some of which are counterintuitive in nature. For example, if $\mu_1 = \ldots = \mu_{k-1} = \mu_k - \delta^*$ and $\sigma_1^2 = \ldots = \sigma_{k-1}^2 > \sigma_k^2$ (where $\sigma^2_i$ is the variance associated with the population with mean $\mu_i$) then there exists a $\delta^*$ close to zero such that $P[CS|\mathcal{P}_R] < 1/k$. It is also shown that additional sampling may cause a decrease in $P[CS|\mathcal{P}_R]$.

A comparison is made between $\mathcal{P}_R$ with $h$ determined by (1.3.131) and $\mathcal{P}_E$. It is shown that $\mathcal{P}_R$ is not always uniformly better than $\mathcal{P}_E$. Expected sample sizes are considered and it is shown that $E(N_1^R) + 1 \geq E(N^E)$ where $N^R$ is the sample size for $\pi_1$ required
by \( P_R \) and \( N_i^E \) is the sample size required by \( P_E \). It is shown that the \( h \) defined by (1.3.131) for \( P_R \) is larger than the \( h \) determined by (1.3.134) for \( P_E \) with equality holding only for \( k = 2 \).

A comparison of \( P[CS] \) was made between \( P_R \) and \( P_E \) under the least-favorable configuration. It is shown that for \( i = 1, 2, \ldots, k \)

\[
\lim_{\sigma_i \to 0} P[CS|P_R] = 1
\]

and

\[
\lim_{n_0 \to \infty} P[CS|P_R] = 1.
\]

Since \( P[CS|P_E]^* = P \) for all \( \sigma_i^2 \) and \( n_0 \) it follows that for large \( n_0 \) or small \( \sigma_i^2 \), \( P_R \) will have larger \( P[CS] \) than \( P_E \) for any configuration of the rest of the parameters. However, it should be noted that \( P_E \) is designed to give exactly what is asked of it regardless of the configuration of the variances.

Tamhane (1975) considers two-stage sampling procedures for selection of the population with the largest mean when the variances are equal but unknown, and also uses the first-stage to eliminate non-contending populations.

Let \( \pi_1, \ldots, \pi_k \) be \( k \geq 2 \) populations such that \( \pi_i \sim N(\mu_i, \sigma^2) \) with \( \mu_i \) and \( \sigma^2 \) unknown. It is desired to select the population with the largest \( \mu_i \). The two-stage sampling procedure \( \mathcal{P}(n_0, 0; s^2, (\delta^*/h) \sqrt{2}) \) used for each population where \( s^2 \) is the usual pooled estimate of \( \sigma^2 \) based on the first \( kn_0 \) observations. Let \( \bar{x}_i^{(1)} \) be the sample mean from \( \pi_i \) after the first-stage sample. Let \( I \) be the subset of populations
where \( k \) and \( \lambda \) are defined by (1.3.135) and (1.3.136). If \( I \) consists of exactly one population, stop sampling and assert that the population in \( I \) is best. If \( I \) consists of two or more populations, choose \( N \) additional observations from those populations.

Now compute the overall sample means of these populations and choose as best that population with the largest sample mean. It is shown that if \( h \) and \( \lambda \) are chosen to satisfy

\[
\begin{align*}
t_k(n_0-1), k-l(h, \ldots, h; \{1/2\}) &= \beta_1 \quad (1.3.135) \\
t_k(n_0-1), k-l(\lambda, \ldots, \lambda; \{1/2\}) &= \beta_2 \quad (1.3.136)
\end{align*}
\]

where \( t_{r,p}(\cdot, \ldots, \cdot; \{\mu\}) \) is the CDF of a \( p \)-variate equicorrelated central \( t \)-distribution with \( r \) degrees of freedom and common correlation \( \rho \) and \( \beta_1, \beta_2 \) are preassigned constants such that

\[
P^* < \beta_1 \leq \beta_2 < 1
\]

and

\[
\beta_1 + \beta_2 - 1 = P^*
\]

then this procedure satisfies the probability requirement. An expression for the expected total sample-size is given and a three-stage screening procedure is considered.

Currently two-stage procedures for estimation of ordered parameters are being considered by H. Chen and should appear in the literature in the near future.
CHAPTER II

ANOVA

2.1. ANOVA in an r-Way Layout

We now consider the r-way layout for the general analysis of variance model. Let $i_1, ..., i_r$ be r subscripts and let

$$X_{i_1i_2...i_rn} = \mu + \lambda_{i_1} + \beta_{i_2} + ... + \gamma_{i_r} + (\lambda\beta)_{i_1i_2} + ... + (\lambda\beta\gamma)_{i_1i_2...i_r} + ... + (\lambda\beta...\gamma)_{i_1i_2...i_r} + e_{i_1i_2...i_rn} \quad (2.1.1)$$

$(i_1 = 1, 2, ..., I_1; \ldots; i_r = 1, 2, ..., I_r, n = 1, 2, \ldots)$ and \{${e_{i_1i_2...i_rn}}$\} are independent random variables such that

$$e_{i_1i_2...i_rn} \sim N(0, \sigma^2_{i_1i_2...i_r}) \quad (2.1.2)$$

The quantities $\lambda_{i_1}, \beta_{i_2}, ..., \gamma_{i_r}$ denote the usual main effects; $(\lambda\beta)_{i_1i_2}, ..., (\lambda\gamma)_{i_1i_r}$ two-factor interactions; and so on, with $(\lambda\beta...\gamma)_{i_1i_2...i_r}$ denoting the r-way interaction. As in the usual r-way layout in analysis of variance (i.e. when error variances are
assumed equal) we impose the side conditions

\[ \sum_{i_1=1}^{I_1} \lambda_{i_1} = 0, \ldots, \sum_{i_r=1}^{I_r} \gamma_{i_r} = 0, \quad (2.1.3) \]

and for any interaction term the sum over one subscript holding the remaining fixed is zero.

We develop tests of the hypotheses

\[ H_1 : \lambda_{i_1} = 0 \quad i_1 = 1, \ldots, I_1 \]
\[ H_2 : \beta_{i_2} = 0 \quad i_2 = 1, \ldots, I_2 \]
\[ \vdots \]
\[ H_r : \gamma_{i_r} = 0 \quad i_r = 1, \ldots, I_r \]
\[ \vdots \]
\[ H_{12 \ldots r} : (\lambda \beta \ldots \gamma)_{i_1 i_2 \ldots i_r} = 0 \quad \text{for all } i_j, j = 1, 2, \ldots, r \]

with test statistics which are independent of the unknown variances, the tests yielding specified power at a given alternative.
Procedure $\mathcal{P}_{B_1}(r)$

In this $r$-way layout there are $I_1 \times I_2 \times \cdots \times I_r$ possible treatment combinations. $\text{Cell}(i_1,i_2,\ldots,i_r)$ refers to the combination of level $i_1$ of the first factor, $i_2$ of the second factor, etc., and level $i_r$ of the $r$th factor. The solution to the testing problem depends upon an application of the two-stage sampling scheme defined by (1.2.23) to each of the $I_1 \times I_2 \times \cdots \times I_r$ cells. That is we apply the two-stage procedure $\mathcal{P}(n_0,1,s_{i_1i_2\ldots i_r}^2)$ to cell$(i_1,i_2,\ldots,i_r)$ where $s_{i_1i_2\ldots i_r}^2$ is the usual unbiased estimate of $s_{i_1i_2\ldots i_r}^2$ based upon the first $n_0$ observations. Let $N_{i_1i_2\ldots i_r}$ be the final sample-size for cell$(i_1,i_2,\ldots,i_r)$. Then we select constants $a_{i_1i_2\ldots i_r} \cdots a_{i_1i_2\ldots i_r}N_{i_1i_2\ldots i_r}$ such that

$$a_{i_1i_2\ldots i_r} = \cdots = a_{i_1i_2\ldots i_r}n_0,$$  \hspace{1cm} (2.1.5a)

$$N_{i_1i_2\ldots i_r} s_{i_1i_2\ldots i_r}^2 \sum_{k=1}^{2} a_{i_1i_2\ldots i_r}^2 = z,$$ \hspace{1cm} (2.1.5b)

$$N_{i_1i_2\ldots i_r} \sum_{k=1}^{r} a_{i_1i_2\ldots i_r}^2 = 1.$$ \hspace{1cm} (2.1.5c)

Compute

$$\bar{x}_{i_1i_2\ldots i_r} = \frac{N_{i_1i_2\ldots i_r}}{\sum_{k=1}^{r} a_{i_1i_2\ldots i_r}^2} x_{i_1i_2\ldots i_r} x_{i_1i_2\ldots i_r} x_{i_1i_2\ldots i_r}$$ \hspace{1cm} (2.1.6)
where \( X_{i_1i_2...i_r} \) is the final set of observations for cell(i_1,i_2,...,i_r). Then let

\[
\bar{x}_{i_1...i_m} = \frac{1}{I_1...I_{m-1}I_{m+1}...I_r} \sum_{i_1=1}^{I_1} \sum_{i_{m-1}=1}^{I_{m-1}} \sum_{i_{m+1}=1}^{I_{m+1}} \sum_{i_r=1}^{I_r} \bar{x}_{i_1i_2...i_r}
\]

\( (m = 1, ..., r) \) \hspace{1cm} (2.1.7)

\[
\bar{x}_{i_1...i_m...i_p} = \frac{1}{I_1...I_{m-1}I_{m+1}...I_{p-1}I_{p+1}...I_r} \\
\sum_{i_1=1}^{I_1} \sum_{i_{m-1}=1}^{I_{m-1}} \sum_{i_{m+1}=1}^{I_{m+1}} \sum_{i_{p-1}=1}^{I_{p-1}} \sum_{i_{p+1}=1}^{I_{p+1}} \sum_{i_r=1}^{I_r} \bar{x}_{i_1i_2...i_r}
\]

\( (m, p = 1, ..., r) \) \hspace{1cm} (2.1.8)

and finally

\[
\bar{x}_{i_1...i_r} = \frac{1}{I_1I_2...I_r} \sum_{i_1=1}^{I_1} \sum_{i_r=1}^{I_r} \bar{x}_{i_1i_2...i_r}
\]

\( (2.1.9) \)

Our choice of test statistics for the hypotheses in (2.1.4) is motivated by an interpretation of the design constant \( z \) (this interpretation is discussed in detail in Section 4.2). Briefly, one
may view $z$ as playing the role of $\sigma^2/N$ in the $r$-way layout when the
events have equal variance $\sigma^2$, $\sigma^2$ is known, and $N$ observations are
taken in each cell. The test statistic one would use in that case for
each of the hypotheses in (2.1.4) is a function of the cell sample
means and $\sigma^2/N$ only. The test statistics we choose have exactly
the same form with (2.1.6) replacing the cell sample means and $z$ re-
placing $\sigma^2/N$.

These test statistics may be generated along the lines of page
124 of Scheffe (1959). In the usual analysis of variance model the
total sum of squares may be partitioned into $2^r+1$ component sums of
squares. Among these there is exactly one corresponding to each main
effect and each of the possible interaction terms. Each when divided
by $\sigma^2$ follows a noncentral chi-square distribution (degrees of freedom
and noncentrality parameter depending upon which main effect or inter-
action is under consideration). Each of these sum of squares (SS)
divided by $\sigma^2$ ($SS/\sigma^2$) is a function of the cell means and $\sigma^2/N$ only.
To obtain the test statistic to be used in the corresponding two-stage
procedure replace the cell means by the generalized cell means (2.1.6)
and $\sigma^2/N$ by $z$ in $SS/\sigma^2$ and use that statistic to test the corre-
sponding hypothesis. For example if we want to test $H_m$, from page 124
of Scheffe (1959)

\[
\frac{SS_m}{\sigma^2} = I_1 I_2 \cdots I_{m-1} I_{m+1} \cdots I_r N \sum_{i_m=1}^{I_m} \frac{(\bar{X}_{i_m} - \bar{X}_{\cdot\cdot\cdot})^2}{\sigma^2}
\]

and so our test statistic for $H_m$ is
To test the hypothesis $H_{mn}$ (i.e. that all two-factor interactions for factors $m$ and $n$ are zero) we use

$$\tilde{F}_m = I_1 I_2 \ldots I_{m-1} I_{m+1} \ldots I_r \sum_{i_m = 1}^{I_m} \left( \tilde{x}_{i_m \ldots} - \tilde{x}_{\ldots} \right)^2$$

Finally, as an example of higher-way interaction tests, suppose we want to test

$$H_{12r} : (\lambda \beta \gamma)_{i_1 i_2 i_r} = 0 \text{ for all } i_1, i_2, i_r$$

and $(\lambda \beta \gamma)_{i_1 i_2 i_r}$ is the usual least squares estimate of this three-way interaction term for factors one, two, and $r$. Then from page 124 of Scheffe's

$$\frac{SS_{12r}}{\sigma^2} = I_3 \ldots I_{r-1} \sum_{i_1 = 1}^{I_1} \sum_{i_2 = 1}^{I_2} \sum_{i_r = 1}^{I_r} \frac{\left[ (\lambda \beta \gamma)_{i_1 i_2 i_r} \right]^2}{\sigma^2}$$

and we take as our test statistic
\[ \tilde{F}_{1,2,r} = I_3 \cdots I_{r-1} \frac{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \sum_{i_r=1}^{I_r} \left[ (\lambda \delta r)_{i_1 i_2 i_r}^* \right]^2}{z} \]  

(2.1.12)

where \((\lambda \delta r)_{i_1 i_2 i_r}^*\) is the same function of the generalized cell means as \((\lambda \delta r)_{i_1 i_2 i_r}\) is of the cell means. (The method just outlined, of choosing a test statistic for any heteroscedastic ANOVA problem is a special case of the Heteroscedastic Method (4.2.12).)

Lemma (2.1.13). \(\{\tilde{X}_{i_1 i_2 \ldots i_r}\} (i_\ell = 1, \ldots, I_\ell, \ell = 1, \ldots, r)\) are independent random variables with

\[
\frac{(\tilde{X}_{i_1 i_2 \ldots i_r} - (\mu + \lambda_{i_1} + \beta_{i_2} + \cdots + (\lambda \delta \ldots r)_{i_1 i_2 \ldots i_r}))}{\sqrt{z}}
\]

distributed as \(t_{n_0-1}\).

Proof. The independence follows from the fact that independent random samples were taken from each cell. The distribution follows as in Stein (1945), page 245.

Lemma (2.1.14). The distributions of the test statistics for hypotheses \(H_1, \ldots, H_{12 \ldots r}\) are independent of the unknown variances.

Proof. From lemma (2.1.13) it follows that each of these statistics is a function of the \(\frac{(\tilde{X}_{i_1 i_2 \ldots i_r})}{\sqrt{z}}\) only which are distributed as
\[ t_{i_1i_2\ldots i_r} + (\mu + \lambda_{i_1} + \beta_{i_2} + \cdots + (\lambda^\beta\ldots r)_{i_1i_2\ldots i_r})/\sqrt{z} \]

where \{t_{i_1i_2\ldots i_r}\}_{i_1 = 1, \ldots, I; i_{\ell} = 1, \ldots, r}\) are i.i.d. \(t_{n_0-1}\) variates. Hence the distributions are independent of the unknown variances.

In general the statistics used to test hypotheses \(H_1, H_2, \ldots, H_{12\ldots r}\) are distributed as specific functions of independent Student's-t variates with \(n_0-1\) degrees of freedom. In the case of the tests for main effects we will now be fully explicit.

**Lemma (2.1.15).** \(\tilde{F}_m\) is distributed as \(I_{1\ldots I_{m-1}I_{m+1}\ldots I_r}(\tilde{t}'B\tilde{t})\) where \(\tilde{t} = (\tilde{t}_1, \ldots, \tilde{t}_m)'\), the \(\{\tilde{t}_i\}_{i=1}^m\) are independent random variables, \(\tilde{t}_i\) is distributed as the average of \(I_1\ldots I_{m-1}I_m\ldots I_r\) i.i.d. Student's-t variates with \(n_0-1\) degrees of freedom plus \((\mu + \eta_i)/\sqrt{z}\), \(\eta_i\) is the main effect of the \(i\)th level of the \(m\)th factor, \(B\) is the \(I_m \times I_m\) matrix \(B = I - 1/I_m 11'\) (I is the \(I_m \times I_m\) identity and \(1\) is the \(I_m\)-dimensional vector of ones) \(m = 1, \ldots, r\).

**Proof.**

\[
\begin{align*}
\bar{x}_{i_1\ldots i_m} &= \frac{1}{I_1\ldots I_{m-1}I_{m+1}\ldots I_r} \\
&\times \sum_{i_1=1}^{I_1} \cdots \sum_{i_{m-1}=1}^{I_{m-1}} \sum_{i_{m+1}=1}^{I_{m+1}} \cdots \sum_{i_r=1}^{I_r} \frac{x_{i_1i_2\ldots i_r}}{\sqrt{z}}
\end{align*}
\]
but
\[ \tilde{X}_{11 \ldots i_r} \sim \frac{t_{11 \ldots i_r}}{\sqrt{z}} \sim t_{11 \ldots i_r} + (\mu + \lambda_1 + \beta_1 + \cdots + (\lambda_\beta \ldots \gamma)_{1i_r}) / \sqrt{z} \]

where \((t_{1i_r})\) are i.i.d. Student's-t variates with \(n_0 - 1\) degrees of freedom, since by (2.1.3)
\[ \sum_{i=1}^{I_1} \lambda_{i1} = 0, \ldots, \sum_{i=1}^{I_r} \lambda_{i1} = 0, \ldots, \sum_{i=1}^{I_1} (\lambda_\beta \gamma)_{1i_r} = 0, \]
and so on,

\[ \tilde{X}_{1m} \ldots / \sqrt{z} \]

Hence \(\tilde{X}_{1m} \ldots / \sqrt{z}\) is distributed as the average of \(I_{1 \ldots i_r}\) independent Student's-t variates plus \((\mu + \eta_{i1}) / \sqrt{z}\). The result now follows from the definition of the matrix \(B\) since \(\bar{t}'B\bar{t} = \sum_{i=1}^{I_m} (\bar{t}_i - \bar{t})\)
where \(\bar{t} = 1 / \sum_{i=1}^{I_m} \bar{t}_i\).

The test for \(H_m\) proceeds by rejecting \(H_m\) if and only if
\[ \tilde{F}_m \geq F_{\alpha, n_0} \] where \(F_{\alpha, n_0}\) is the upper \(\alpha\)th percent point of the null distribution of \(\tilde{F}_m\). Tests for the other hypotheses are performed in a similar manner, namely reject the hypothesis \(H_1, \ldots, H_r, \ldots,\) or
$H_{12...r}$ if and only if the corresponding statistic $\tilde{F}_1, \ldots, \tilde{F}_r, \ldots$, or $\tilde{F}_{12...r}$ is greater than the upper $\alpha$th percent point of their respective null distributions (see Heteroscedastic Method 4.2.12).

Note that the general linear model under conditions of heteroscedasticity may be considered using the Heteroscedastic Method of Chapter 4.

**Limiting Distributions**

As noted above the exact distributions of our test statistics depend upon specific functions of independent Student's-t variates. These distributions are not tabulated in the literature yet (but, see Chapter 3 for some extensive tables), and thus an approximation is called for. We consider approximating the distributions of these test statistics by their limiting distributions as $n_0$, the initial sample size, approaches infinity.

To obtain these limiting distributions the following lemmas will be needed.

**Lemma (2.1.16).** Let $T : \mathbb{R}^k \to \mathbb{R}^m$ be a continuous mapping and let $X_1, X_2, \ldots$ be a sequence of $k$-dimensional random variables such that $X_n$ converges in distribution to $X$. Then $T(X_n)$ converges in distribution to $T(X)$. 
Proof. See page 90 of Tucker (1967).

Lemma (2.1.17). If $Y$ is distributed as $N_p(w, V)$, then $Y'BY \sim \chi^2(\Delta)$, $\Delta = w'By$ (where $k$ is the rank of $B$) if and only if $BV$ is idempotent.

Proof. See page 84 of Graybill.

Theorem (2.1.18). The limiting distribution (as $n_0$ approaches infinity) of the test statistic $\widetilde{F}_m$ is $\chi^2_{m-1}(\Delta)$, where $\Delta = I_1 \cdots I_{m-1} I_{m+1} \cdots I_r$ and

$$\sum_{i=1}^{m} \eta^2_{i m} / n (m = 1, 2, \ldots, r).$$

Proof. From Lemma (2.1.15) $\widetilde{F}_m$ is distributed as a continuous function of the independent random variables $t_{i m}^+ (w + \eta_{i m}) / \sqrt{z}$ which converge in distribution to $N(\mu + \eta_{i m}) / \sqrt{z} I_1 \cdots I_{m-1} I_{m+1} \cdots I_r$ since $t_{i m}$ is the average of $I_1 \cdots I_{m-1} I_{m+1} \cdots I_r$ i.i.d. Student's-t variates with $n_0-1$ degrees of freedom. Now $\widetilde{F}_m \sim I_1 \cdots I_{m-1} I_{m+1} \cdots I_r (t_1, \ldots, t_{m r}) B(t_1, \ldots, t_{m r})'$ where $B$ is the matrix defined in Lemma (2.1.15), which is idempotent and of rank $I_{m-1}$. The result now follows from Lemma (2.1.16) and Lemma (2.1.17).

Let $(\eta^r)_{i m n}$ denote the two-factor interaction between factors $m$ and $n$. As we have described we would use the statistic $\widetilde{F}_{m,n}$ to test the hypothesis that each of these two-factor interactions are zero. The limiting distribution of this statistic is given by the following theorem.
Theorem (2.1.19). The limiting distribution of the test statistic
\[ \tilde{F}_{m,n} \] is a noncentral chi-square with \((I_m-1)(I_n-1)\) degrees of freedom
and noncentrality parameter \(K \Sigma \sum_{i} (\eta \tau)^2\), where \(K = I_1 \ldots I_{m-1} I_{m+1} \ldots I_{n-1} I_{n+1} \ldots I_r / \sigma^2\) \((m,n = 1, \ldots, r)\).

Proof. \(\tilde{F}_{m,n}\) is a continuous function of the \(\widetilde{X}_{i_1 i_2 \ldots i_r}\) which converge in distribution to
\[ Z_{i_1 i_2 \ldots i_r} \sim N(\mu + \lambda_{i_1} + \cdots + \gamma_{i_r} + \cdots + (\eta \tau)_{i_m i_n} + \cdots + (\lambda \beta \cdots \gamma)_{i_1 i_2 \ldots i_r} \sigma^2) \]

Thus by lemma (2.1.16) \(\tilde{F}_{m,n}\) is distributed in the limit as the same function with the \(\widetilde{X}_{i_1 i_2 \ldots i_r}\) replaced by \(Z_{i_1 i_2 \ldots i_r}\). From the usual (see Scheffe' (1959), page 124) least square theory that function is distributed as a noncentral chi-square with \((I_m-1)(I_n-1)\) degrees of freedom and noncentrality parameter \(K \Sigma \sum_{i} (\eta \tau)^2\) and the results follow.

The limiting distributions of the remaining test statistics are obtained in a similar manner. As shown the test statistics are functions of \(\widetilde{X}_{i_1 i_2 \ldots i_m}\) which converge a distribution to \(Z_{i_1 i_2 \ldots i_m}\)
\[ \sim N(\mu + \lambda_{i_1} + \cdots + \gamma_{i_r} + \cdots + (\eta \tau)_{i_m i_n} + \cdots + (\lambda \beta \cdots \gamma)_{i_1 i_2 \ldots i_r} \sigma^2) \]
The limiting distributions are obtained by replacing the \(\widetilde{X}_{i_1 i_2 \ldots i_r}\) with \(Z_{i_1 i_2 \ldots i_r}\).

As an example of these limiting distributions consider the following. Let \(I_1 = 4, \lambda_1 = 1, \lambda_2 = -2, \lambda_3 = 1\) and \(\lambda_4 = 0\). Then \(\tilde{F}_1\) is distributed in the limit as \(\chi^2_3(6/z)\). Thus to make a test at level \(\alpha = .05\) we reject \(H_1\) if and only if \(\tilde{F} > \chi^2_3(.05)\). To achieve .95 at the alternative \(\sum_{i=1}^4 \lambda_i^2 = 6\) we need to choose \(z\) so that \(6/z = 17.17\) which is the noncentrality required, hence \(z = .349\).
2.2. Multiple-Comparison Procedures

As in the case of usual F-tests in the r-way layout, if we reject some of the hypothesis \((H_1, H_2, \ldots, H_{12\ldots r})\) we will usually want to investigate the cause of the rejecting further (note that if the multiple comparisons are done only when the experimenter rejects the null hypothesis the confidence coefficients change). In this section we consider multiple-comparison procedures analogous to those of Tukey and Scheffé for contrasts and linear combinations of the parameters in the r-way model. Those multiple-comparison procedures yield families of simultaneous confidence intervals of fixed length with confidence coefficient independent of the unknown variances.

2.2.A. All Pairwise Differences

We first consider families of simultaneous confidence intervals for all \(\binom{I_m}{2}\) differences \((\eta_{i_m} - \eta_{i_m'})\) of main effects for factor \(m\) \((m = 1, \ldots, r)\).

**Theorem (2.2.1).** The probability is \(1-\alpha\) that all \(\binom{I_m}{2}\) differences \((\eta_{i_m} - \eta_{i_m'})\) are simultaneously covered by the family of intervals

\[
\left[\bar{X}_{i_m} - \bar{X}_{i_m'} \pm \sqrt{\frac{q^\alpha}{I_m}} \sqrt{\frac{1}{\eta_0 - 1}}\right]
\]

where \(q^\alpha_{I_m, \eta_0 - 1}\) is the upper \(\alpha\)th percent point of the range \(R_m\) of \(I_m\) independent random variables \([Z_i]_{i=1}^m\) where \(Z_i\) is distributed as the average of \(I_1I_{m-1}I_{m+1}\ldots I_r\) independent Student's-t variates with \(\eta_0 - 1\) degrees of freedom.
Proof. Note that
\[
\tilde{X}_{i_1 \ldots i_m} - (\mu + \eta_{i_m}) \sim \frac{Z_{i_m}}{\sqrt{v}}, \quad i_m = 1, \ldots, I_m.
\]

Now
\[
\left| \tilde{X}_{i_1 \ldots i_m} - (\mu + \eta_{i_m}) \right| - \left| \tilde{X}_{i_1 \ldots i_m} - (\mu + \eta_{i_m}) \right| \leq \frac{\alpha}{\sqrt{v}} I_m, n_0 - 1
\]
for all \(i_m, i_m' = 1, 2, \ldots, I_m\) if and only if \(R_{I_m} \leq \frac{\alpha}{\sqrt{v}} I_m, n_0 - 1\), and the result follows.

2.2.B. Contrasts

We denote the space of all contrasts for the \(m\)th factor by
\[
\mathcal{L} = \left\{ \mathbf{c} = (c_1, \ldots, c_{I_m}) / \sum_{i_m=1}^{I_m} c_{i_m} = 0 \right\}.
\]

Lemma (2.2.2). If \(|Y_{i} - Y_{j}| \leq c'\) for all \(i, j = 1, \ldots, I_m\) then
\[
|\sum_{i=1}^{I_m} c_i Y_i | \leq c' (\sum_{i=1}^{I_m} |c_i|) / 2 \quad \text{for all } \mathbf{c} \in \mathcal{L}_{c_m}.
\]


Theorem (2.2.3). The probability is \(1-\alpha\) that all contrasts \(\sum_{i_m=1}^{I_m} c_{i_m} \eta_{i_m}\) are simultaneously covered by the family of intervals
Proof. We need only show that the events

\[
A : \left\{ \sum_{i_m=1}^{I_m} c_{i_m} \tilde{x}_{i_m} \cdots \pm \frac{1}{2} \sum_{i_m=1}^{I_m} |c_{i_m}| \text{ for all } c \in \mathcal{L}_m \right\}
\]

and

\[
B : \left\{ \sum_{i_m=1}^{I_m} c_{i_m} \tilde{x}_{i_m} \cdots - \left( \sum_{i_m=1}^{I_m} c_{i_m} \tilde{x}_{i_m} \cdots \right) \text{ for all } i_m, i_m' = 1, \ldots, I_m \right\}
\]

are equivalent. Clearly the appropriate choice of \( c_{i_m} \) yields \( A \Rightarrow B \), from Lemma (2.2.2) \( B \Rightarrow A \), and the result follows.

2.2.C. All Linear Combinations: Augmented Range

We next construct a family of fixed-width confidence intervals for all linear combinations of the main effects of the \( m \)th factor. The procedure is based on the augmented range of the variables \( Z_{i_m} \) defined in Theorem (2.2.1).

Definition (2.2.4). The augmented range of the variables \( Z_{i_m} \) is defined by

\[
Q_{I_m, n_0} = \max\{ |M|_m, R_{I_m} \}
\]

where

\[
|M|_m = \max_{1 \leq i_m \leq I_m} |Z_{i_m}|.
\]
A random variable with the same distribution as $Q_{I_m, n_0}$ is defined by

$$Q_{I_m, n_0}^i = \max_{i, j=0,1, \ldots, I_m} |Z_i - Z_j|$$

where $Z_i$, $i = 1, \ldots, I_m$ are defined as in Theorem (2.2.1) and are independent of $Z_0 \equiv 0$.

**Definition (2.2.5).** Let $\ell = (\ell_1, \ldots, \ell_{I_m}) \in \mathbb{R}^m$, then $\sum_{i=1}^{I_m} \ell_i \eta_{i_i}$ is a linear combination of the main effects $\{\eta_{i_i}\}$ for factor $m$.

Let $P = \{i_{m}/\ell_{i_m} > 0\}$ and $N = \{i_{m}/\ell_{i_m} < 0\}$ and define the function $M(\ell_1, \ldots, \ell_{I_m})$ by

$$M(\ell_1, \ldots, \ell_{I_m}) = \max \{ \sum_{i \in P} \ell_i, - \sum_{i \in N} \ell_i \} \quad (2.2.6)$$

Let $q_{I_m, n_0}^{\alpha}$ be the upper $\alpha$th percent point of the distribution of $Q_{I_m, n_0}$, then we have the following theorem.

**Theorem (2.2.7).** The probability is $1-\alpha$ that all linear combinations $\sum_{i=1}^{I_m} \ell_i \eta_{i_i}$ are simultaneously covered by the family of intervals

$$\{ \sum_{i=1}^{I_m} \ell_i \eta_{i_i} \ldots \pm \sqrt{z} q_{I_m, n_0}^{\alpha} M(\ell_1, \ldots, \ell_{I_m}) \text{ for all } \ell \in \mathbb{R}^m \}.$$
Proof. Taking $\tilde{X}_0 = 0$ and $\eta_0 = 0$ where $\tilde{X}_0$ is independent of $(\tilde{X}_{i_1}, \ldots, \tilde{X}_{i_m})$, the linear combination $\sum_{i_m=1}^{I_m} \xi_{i_m} \tilde{X}_{i_m} \ldots$ can be written as the contrast $\sum_{i_m=0}^{I_m} c_{i_m} \tilde{X}_{i_m} \ldots$, where $c_{i_m} = \xi_{i_m}$ for $i_m = 1, \ldots, I_m$ and $c_0 = -\sum_{i_m=1}^{I_m} \xi_{i_m}$. From Lemma (2.2.2) and by the appropriate choices of the $c_{i_m}$ (i.e. $c_{i_m} = +1$, $c_{i_m} = -1$), the two events

$$\max_{i_m, i'_m = 0, \ldots, I_m} \left\{ |\tilde{X}_{i_m} \ldots - \tilde{X}_{i'_m} \ldots - (\eta_{i_m} - \eta_{i'_m})| \right\} \leq \sqrt{c'}$$

and

$$\left| \sum_{i_m=0}^{I_m} c_{i_m} (\tilde{X}_{i_m} \ldots - \eta_{i_m} \ldots) \right| \leq \sqrt{c'} \sum_{i_m=0}^{I_m} c_{i_m}$$

are equivalent. Now the distribution of

$$\max_{i_m, i'_m = 0, \ldots, I_m} \left( \frac{|\tilde{X}_{i_m} \ldots - \tilde{X}_{i'_m} \ldots - (\eta_{i_m} - \eta_{i'_m})|}{\sqrt{c'}} \right)$$

is the same as $Q_{I_m, \eta_0}$. From the definition of the $c_{i_m}$'s we have

$$\sum_{i_m=0}^{I_m} |c_{i_m}| = |c_0| + \sum_{i_m=1}^{I_m} |c_{i_m}| = |\sum_{i_m=1}^{I_m} \xi_{i_m}| + \sum_{i_m=1}^{I_m} |\xi_{i_m}| = |\sum_{i_m=1}^{I_m} \xi_{i_m}| + \sum_{i_m=1}^{I_m} |\xi_{i_m}| = |\sum_{i_m=1}^{I_m} \xi_{i_m}| + \sum_{i_m=1}^{I_m} \xi_{i_m} - \sum_{i_m=1}^{I_m} \xi_{i_m}$$

If $\sum_{i_m \in P} \xi_{i_m} > -\sum_{i_m \in N} \xi_{i_m}$, then $M(\ell_{1m}, \ldots, \ell_{Im}) = \sum_{i_m \in P} \ell_{i_m}$ and
\[
\sum_{i_m=0}^{I_m} |c_{i_m}| = \sum_{i_m \in P} \ell_{i_m} + \sum_{i_m \in N} \ell_{i_m} + \sum_{i_m \in P} \ell_{i_m} - \sum_{i_m \in N} \ell_{i_m}
\]

\[
= 2 \sum_{i_m \in P} \ell_{i_m} = 2(\ell_1, \ldots, \ell_{I_m})
\]

Similarly if \(\sum_{i_m \in N} \ell_{i_m} < -\sum_{i_m \in P} \ell_{i_m}\) then

\[
\sum_{i_m=0}^{I_m} |c_{i_m}| = 2M(\ell_1, \ldots, \ell_{I_m})
\]

Hence

\[
\frac{1}{2} \sum_{i_m=0}^{I_m} |c_{i_m}| = M(\ell_1, \ldots, \ell_{I_m})
\]

and if we choose \(c' = q_{I_m, n_0}^{\alpha}\) the theorem follows since

\[
P[\left| \sum_{i_m=0}^{I_m} c_{i_m} (\bar{x}_{i_m} - \bar{\eta}_{i_m} \ldots) \right| \leq \sqrt{2} q_{I_m, n_0}^{\alpha} M(\ell_1, \ldots, \ell_{I_m})] = P[Q_{I_m, n_0} \leq q_{I_m, n_0}^{\alpha}] = 1-\alpha
\]

2.2.D. All Linear Combinations: Scheffe Type

We next consider a family of simultaneous confidence intervals for all linear combinations. We consider this family to be in the spirit of Scheffe's simultaneous intervals because its validity depends upon the Cauchy-Schwarz inequality in the same manner as does Scheffe's.
Lemma (2.2.8). For $c > 0$, $|\sum_{i=1}^{k} a_i y_i| \leq c(\sum_{i=1}^{k} a_i^2)^{1/2}$ for all $(a_1, \ldots, a_k)$ if and only if $\sum_{i=1}^{k} y_i^2 \leq c^2$.


Theorem (2.2.9). The probability is $1-\alpha$ that all linear combinations
$$\sum_{i=1}^{m} l_i \eta_i$$
are simultaneously covered by the family of intervals
$$\left\{ \sum_{i=1}^{m} l_i (\tilde{X}_i - \bar{X}_i) \leq \sqrt{z} \frac{1}{m} \xi_i^{1/2} \right\}$$
for all $(\tilde{l}_1, \ldots, \tilde{l}_m)$ if and only if
$$\sum_{i=1}^{m} \left( \tilde{X}_i - \bar{X}_i \right)^2 \leq c^2 \frac{1}{m} \xi_i^{1/2}.$$

where $\xi_i^{1/2}$ is the upper $\alpha$th percent point for the distribution of
$$Q = (Z_1, \ldots, Z_m) B(Z_1, \ldots, Z_m)'$$
and $B$ is defined as in Lemma (2.1.15).

Proof. From Lemma (2.2.8)
$$\left| \sum_{i=1}^{m} \frac{\tilde{X}_i - \bar{X}_i}{\sqrt{z}} \eta_i \right| \leq c(\sum_{i=1}^{m} \xi_i^{1/2})^{1/2}$$
for all $(\tilde{l}_1, \ldots, \tilde{l}_m)$ if and only if
$$\sum_{i=1}^{m} \frac{\tilde{X}_i - \bar{X}_i}{\sqrt{z}} \eta_i^2 \leq c^2.$$

Since $[(\tilde{X}_i - \bar{X}_i)/(\mu + \eta_i)] \xi_i^{1/2}$ are i.i.d. variates distributed as
$$\sum_{i=1}^{m} \frac{1}{m} [Z_i]' \xi_i^{1/2}$$
and letting $c^2 = \xi_i^{1/2}$ the result follows since
2.2.E. Contrasts: Scheffe' Type

In a manner similar to that which yielded intervals for all linear combinations in Section 2.2.D, we may employ the Cauchy-Schwarz inequality to develop simultaneous confidence intervals for all contrasts among main effects.

Theorem (2.2.10). The probability is at least $1 - \alpha$ that all contrasts

$$\sum_{i=1}^{m} c_i \eta_i$$

are simultaneously covered by the family of intervals

$$\left\{ \sum_{i=1}^{m} c_i \bar{x}_{i,m} \pm \sqrt{z} \left( \xi_{I, n_0 - 1}^2 \right)^{1/2} \left( \sum_{i=1}^{m} c_i^2 \right)^{1/2} \right\}$$

for all \(\{c_1, \ldots, c_i\}\) such that \(\sum_{i=1}^{m} c_i = 0\), where \(\xi_{I, n_0 - 1}^2\) is the upper \(\alpha\)th percent point of the distribution of

$$Q = (Z_1, \ldots, Z_I) B(Z_1, \ldots, Z_I)' \left( \sum_{i=1}^{m} c_i \right)$$

are defined as in Theorem (2.2.1), and \(B\) is defined as in Lemma (2.1.15).
Proof. From Lemma (2.2.8) if we let \( Y_i = Z_i - \bar{Z} \) where \( \bar{Z} = 1/\sum_{i=1}^{m} Z_i \), then

\[
Q = \sum_{i=1}^{m} (Z_i - \bar{Z})^2 \leq \xi_{m,0}^{-1}. \]

if and only if

\[
\left| \sum_{i=1}^{m} a_i Y_i \right| \leq (\xi_{m,0}^{-1})^{1/2} (\sum_{i=1}^{m} a_i^2)^{1/2}
\]

for all \((a_1, \ldots, a_m) \in \mathbb{R}^m\). Thus in particular

\[
Q \leq \xi_{m,0}^{-1} \Rightarrow \left\{ \sum_{i=1}^{m} c_i (Z_i - \bar{Z}) \leq (\xi_{m,0}^{-1})^{1/2} (\sum_{i=1}^{m} c_i^2)^{1/2} \right\}
\]

for all \((c_1, \ldots, c_m)\) such that \(\sum_{i=1}^{m} c_i = 0\)

\[
\Rightarrow \left\{ \sum_{i=1}^{m} c_i Z_i \leq (\xi_{m,0}^{-1})^{1/2} (\sum_{i=1}^{m} c_i^2)^{1/2} \right\}
\]

since \(\sum_{i=1}^{m} c_i = 0\).

Now \((\tilde{\mathcal{X}}_{i-m} \ldots - (\mu + \eta_i))/\sqrt{\bar{Z}} \sim Z_i\), so

\[
[Q \leq \xi_{m,0}^{-1}] \Rightarrow \left\{ \sum_{i=1}^{m} c_i \frac{\tilde{\mathcal{X}}_{i-m} \ldots - \eta_i}{\sqrt{\bar{Z}}} \leq (\xi_{m,0}^{-1})^{1/2} (\sum_{i=1}^{m} c_i^2)^{1/2} \right\}
\]

for all \((c_1, \ldots, c_m)\) such that \(\sum_{i=1}^{m} c_i = 0\)

and the result follows.
Note that similar multiple comparison procedures may be obtained for the interaction effects by following the Heteroscedastic Method which is presented in Section 4.2. However, as Scheffe (1959, p. 109) points out such comparisons usually are not made.

**Approximate Confidence Intervals.**

As with the distributions of the test statistics discussed in Section 2.1 the distributions required for implementation of the above confidence interval procedures are not yet (in most cases) tabled, (although Hochberg (1975) has tables for the range of $k$ i.i.d. $t_{n_0-1}$ variates, and Dudewicz and Dalal (forthcoming) have tables for all pairwise differences, and Dudewicz, Ramberg and Chen for multiple comparisons with a control). Until they are a method of approximation based on tabled distributions is needed. We consider the behaviour of our intervals as $n_0$ approaches infinity. Thus in the case of all $I_m(I_m-1)/2$ differences of $(\eta_i - \eta_i')$, for example, the form of the family of intervals stays the same but $q_{n_0-1}^\alpha$ is replaced by the range of $I_m$ independent random variables $\{Z'_i\}_{i=1}^m$ which are distributed as the average of $I_1...I_{m-1}I_{m+1}...I_r$ independent $N(0,1)$ random variables. That is we can approximate the intervals by

$$\{ \tilde{x}_{i_1} - \tilde{x}_{i_2} \pm \sqrt{z/I_1...I_{m-1}I_{m+1}...I_r} q_{\alpha} \}$$

where $q_{\alpha}$ is the upper $\alpha$th percent point of the range of $I_m$ independent $N(0,1)$ random variables.
Similar approximate intervals can be constructed, using the Heteroscedastic Method of Chapter 4, for the case of all contrasts or all linear combinations in the other procedures by noting that the limiting distribution of $\tilde{X}_{i_1i_2\ldots i_r}$ is $N(\mu + \lambda_{i_1} + \cdots + \gamma_{i_r} + \cdots + (\lambda_{\beta\ldots\gamma})_{i_1i_2\ldots i_r}, z)$ as $n_0$ approaches infinity.

In the next two sections we consider the general procedure $P_{B_r}(r)$ for the important special cases of $r = 2$ and $r = 1$. These two cases are the most often encountered in practice and provide good examples of how the general theory is to be applied.
2.3. ANOVA in a Two-Way Layout

When \( r = 2 \) the general model (2.1.1) reduces to

\[
X_{ij} = \mu + \alpha_i + \beta_j + \alpha_\beta_{ij} + e_{ijk}
\]

(\( i = 1, \ldots, I; j = 1, \ldots, J; k = 1,2, \ldots \)).

Assume that the \( \{e_{ijk}\} \) are independent random variables with

\[
e_{ijk} \sim N(0, \sigma^2_{ij})
\]

and that

\[
\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = \sum_{i=1}^{I} \alpha_\beta_{ij} = \sum_{j=1}^{J} \alpha_\beta_{ij} = 0
\]

The hypotheses under consideration in Section 2.1 then reduce to

\[
H_0 : \alpha_i = 0 \quad \text{for all } i,
\]

(2.3.2)

\[
H_1 : \beta_j = 0 \quad \text{for all } j,
\]

(2.3.3)

\[
H_2 : \alpha_\beta_{ij} = 0 \quad \text{for all } i \text{ and } j.
\]

(2.3.4)

We consider tests of these hypotheses based on test statistics which are independent of the unknown variances.

Based on the general theory of Section 2.1 we apply procedure \( B_1 \) (2) to this problem. Thus we base our tests of (2.3.2) on the test statistic

\[
\bar{F}_0 = \frac{J \sum_{i=1}^{I} \frac{\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot}}{z}^2}{z},
\]

(2.3.5)

hypothesis (2.2.3) is tested using
\[ F_1 = I \sum_{j=1}^{J} \frac{(\bar{x}_{i,j} - \bar{x}_{..})^2}{\bar{z}} , \quad (2.3.6) \]

and hypothesis (2.3.4) is tested using

\[ F_2 = I \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(\bar{x}_{i,j} - \bar{x}_{..} - \bar{x}_{..} + \bar{x}_{..})^2}{\bar{z}} . \quad (2.3.7) \]

The distributions of each of these statistics is independent of the unknown cell variances and each has a limiting noncentral chi-square distribution given by Theorems (2.1.18) and (2.1.19). Thus \( F_0 \) has a limiting \( \chi^2_{I-1}(\Delta_0) \) where \( \Delta_0 = I \sum_{i=1}^{I} \alpha_i^2 / \bar{z} \). \( F_1 \) has a limiting distribution which is \( \chi^2_{J-1}(\Delta_1) \) where \( \Delta_1 = I \sum_{j=1}^{J} \beta_j^2 / \bar{z} \). Similarly the limiting distribution of \( F_2 \) is \( \chi^2_{(I-1)(J-1)}(\Delta_2) \) with \( \Delta_2 = I \sum_{i=1}^{I} \sum_{j=1}^{J} (\alpha \beta_{i,j})^2 / \bar{z} \).

### 2.3.A. Multiple-Comparisons

Multiple-comparison procedures follow directly from the general theory of Section 2.2. For example, if we are interested in all pairwise comparisons of the main effects \( \{\alpha_i\} \) then with probability \( 1-\alpha \) the family of intervals

\[ \{\tilde{x}_{i.} - \tilde{x}_{..} \pm \sqrt{\bar{z} q_{\alpha}^{\alpha}} \} \quad (2.3.8) \]

simultaneously covers all \( I(I-1)/2 \) differences \( (\alpha_i - \alpha_{..}) \), where
$\bar{q}_{\alpha}^\prime$ is the upper $\alpha$th percent point of the range of the $I$ independent random variables $\{Z_i\}_{i=1}^I$. $Z_i$ is distributed as the average of $J$ i.i.d. Student-$t$ variates with $n_0-1$ degrees of freedom. Similar results for all contrasts and all linear combinations follow directly from Theorems (2.2.3), (2.2.7), (2.2.9), and (2.2.10) by taking $r = 2$.

As was discussed, these intervals may be approximated by limiting forms of the intervals when $n_0 \to \infty$. Thus, the intervals (2.3.8) can be approximated by the intervals

$$[\bar{\tilde{x}}_i - \bar{\tilde{x}}_i, \ldots \pm \sqrt{z/\beta} \bar{q}_{\alpha}^\prime],$$

(2.3.9)

where $\bar{q}_{\alpha}^\prime$ is the upper $\alpha$th percent point of the range of $I$ independent $N(0,1)$ random variables. Similar results hold for the other multiple comparisons.
In this section we consider the application of procedure \( P_{B_1} (r) \) to the one-way layout (i.e. the case \( r = 1 \)) and denote the procedure in this case by \( P_{B_1} \). In addition in Sections 2.5 and 2.6 we develop two other test procedures \( P_{B_2} \) and \( P_{B_3} \) which possess the desired property of yielding test statistics, for the hypothesis of equality of means, which have distributions independent of the unknown variances which are not derived from the general theory of Section 2.1.

The one-way layout may be set up as follows: let

\[
X_{ij} = \mu_i + e_{ij} \quad (i = 1, 2, \ldots, k; \ j = 1, 2, \ldots) , \tag{2.4.1}
\]

where the \( \{e_{ij}\} \) are independent random variables with \( e_{ij} \sim N(0, \sigma_i^2) \), \((\infty < \mu_i < \infty, 0 < \sigma_i^2 < \infty, \text{unknown})\). Our goal is to test the null hypothesis

\[
H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \tag{2.4.2}
\]

in such a way that the power of the test is controllable and not dependent upon the unknown variances.

When \( k = 2 \), \( P_{B_1} \) reduces to a form equivalent to Chapman's test (1.3.57) for the Behrens-Fisher problem. For general \( k \), \( P_{B_2} \) is based on a particular quadratic form in generalized means. \( P_{B_3} \) combines Chakravarti's generalization of Scheffé's solution to the Behrens-Fisher problem and Chatterjee's extension of Stein's results. All three procedures are based on the two-stage sampling scheme given by (1.2.23).
Based on the general theory of Section 2.1 procedure $\mathcal{P}_{B_1}$ bases the test of (2.4.2) on the statistic

$$\tilde{F} = \frac{\sum_{i=1}^{k} (\tilde{x}_i - \tilde{x})^2}{z}$$

(2.4.3)

and rejects $H_0 : \mu_1 = \cdots = \mu_k$ if and only if

$$\tilde{F} \geq \tilde{F}_{k,n_0-1}^{\alpha}$$

where $\tilde{F}_{k,n_0-1}^{\alpha}$ is the upper $\alpha$th percent point of the distribution of $Q = (t_1, \ldots, t_k) B(t_1, \ldots, t_k)' (t_1, \ldots, t_k)'$ being a vector of i.i.d. Student's-t variates with $n_0-1$ degrees of freedom. The matrix

$$B = I_k - \frac{1}{k} \; 1' 1$$

(2.4.4)

where $I_k$ is the $k \times k$ identity matrix and $1$ is a $1 \times k$ vector of ones.

Properties of $\tilde{F}$

The general properties of $\tilde{F}$ follow from the theory in Section 2.1. That is the distribution of $\tilde{F}$ is independent of $\sigma_1^2, \ldots, \sigma_k^2$. However, in this case we can be more specific. From lemma (2.1.15) it follows that $\tilde{F}$ is a quadratic form in i.i.d. Student's-t variates. That is since $\tilde{x}_i / \sqrt{z} \sim t_1 + \mu_1 / \sqrt{z}$ we may write
\[ F = \sum_{i=1}^{k} \left( t_i - \bar{t} + \frac{(\mu_i - \bar{\mu})}{\sqrt{z}} \right)^2 \] \hspace{1cm} (2.4.5)

where \( \bar{t} = \frac{1}{k} \sum_{i=1}^{k} t_i \) and \( \bar{\mu} = \frac{1}{k} \sum_{i=1}^{k} \mu_i \). Under \( H_0 : \mu_1 = \cdots = \mu_k \), \( \mu_i - \bar{\mu} = 0 \) for each \( i \) and hence \( F \) is distributed as \( Q \) above.

**Lemma (2.4.6).** The power of the test based on \( \widehat{F} \) converges to one as 
\[ \mu_i - \bar{\mu} \to \pm \infty \] for any \( i = 1, \ldots, k \).

**Proof.**
\[
P[F > F_{k, n_0 - 1}^{\alpha}] \geq P \left[ \left( t_i - \bar{t} + \frac{(\mu_i - \bar{\mu})}{\sqrt{z}} \right)^2 \geq F_{k, n_0 - 1}^{\alpha} \right] \text{ for any } i, \text{ and}
\]

\[
P \left[ \left( t_i - \bar{t} + \frac{(\mu_i - \bar{\mu})}{\sqrt{z}} \right)^2 \geq F_{k, n_0 - 1}^{\alpha} \right] = 1 - P \left[ \sqrt{F_{k, n_0 - 1}^{\alpha}} - \frac{(\mu_i - \bar{\mu})}{\sqrt{z}} \leq t_i - \bar{t} \leq \sqrt{F_{k, n_0 - 1}^{\alpha}} - \frac{(\mu_i - \bar{\mu})}{\sqrt{z}} \right]
\]

hence for any \( i \)

\[
\lim_{(\mu_i - \bar{\mu}) \to \pm \infty} P \left[ \left( t_i - \bar{t} + \frac{(\mu_i - \bar{\mu})}{\sqrt{z}} \right)^2 \geq F_{k, n_0 - 1}^{\alpha} \right] = 1
\]

and the lemma follows.

Note it is easy to show that the random variable \( t_i - \bar{t} \) has a symmetric distribution and thus the lower bound for \( P[F > F_{k, n_0 - 1}^{\alpha}] \) is monotonically increasing as \( (\mu_i - \bar{\mu}) \to \pm \infty \). From this we see that the power is controllable through \( z \) since it follows from Lemma (2.4.6), that, at any fixed alternative,

\[
\lim_{z \to 0} P[F > F_{k, n_0 - 1}^{\alpha}] = 1. \hspace{1cm} (2.4.7)
\]
It follows from Theorem (2.1.18) that the limiting distribution of \( \tilde{F} \) as \( n_0 \to \infty \) is \( \chi^2_{k-1}(\Delta) \) where \( \Delta = \sum_{i=1}^{k} (u_i - \bar{u})^2 / z \). However, in this case we suggest approximating the distribution of \( \tilde{F} \) by that of \( (n_0-1)/(n_0-3) \cdot \chi^2_{k-1} \) random variable. In this case the two random variables have the same expected value and the numerical results of Chapter 3 indicate that this is an excellent approximation.

Hochberg (1975) suggested approximating the null distribution, of \( \tilde{F} \) by that of an \( \chi^2_{\ell, m} \) variable where \( \ell \) and \( m \) are chosen so that \( \tilde{F} \) and \( \chi^2_{\ell, m} \) have the same first two moments. Tables of the upper 10\% and 5\% points for \( k = 3(1)10 \) and \( n_0 = 5, 8, 10, 15, 20(10)60 \) are given there.

### Equivalence of Chapman's Statistic to \( \tilde{F} \) When \( k=2 \).

When \( k = 2 \) we are actually considering the classical Behrens-Fisher problem. For \( k = 2 \) Chapman (1950) provided, almost inter alia, a test of the null hypothesis \( H_0 : \mu_1 = \mu_2 \) which was based on the statistic \( t' = (\tilde{X}_1 - \tilde{X}_2) / \sqrt{z} \). For the case \( k = 2, \tilde{X}_i = \frac{1}{2} (\tilde{X}_1 + \tilde{X}_2) \) and

\[
\tilde{F} = \sum_{i=1}^{2} \frac{(\tilde{X}_i - \tilde{X}_*)^2}{z} = \frac{1}{z} \left[ \left( \tilde{X}_1 - \left( \frac{\tilde{X}_1 + \tilde{X}_2}{2} \right) \right)^2 + \left( \tilde{X}_2 - \left( \frac{\tilde{X}_1 + \tilde{X}_2}{2} \right) \right)^2 \right]
\]

\[
= \frac{1}{z} \left[ \left( \frac{\tilde{X}_1 - \tilde{X}_2}{4} \right)^2 + \left( \frac{\tilde{X}_2 - \tilde{X}_1}{4} \right)^2 \right] = \frac{1}{2z} (\tilde{X}_1 - \tilde{X}_2)^2 = \frac{1}{2} t'^2
\]

which is equivalent to Chapman's test statistic \( t' \). Thus, the test procedure of Chapman is a special case of our procedure.
We also note the analogy here to the case \( k = 2 \) when the variances are equal. That is, when \( k = 2 \) the usual single-sample test statistic \( F \) equals \( t^2/2 \) where \( t \) is the usual \( t \)-statistic which tests the same hypothesis as \( F \). Thus \( \widetilde{F} \) and \( t' \) are related in the same way as their single-stage counterparts \( F \) and \( t \).

**Expected Sample Size**

The total sample size for \( \mathcal{B}_1 \)

\[
N_{T_1} = N_1 + N_2 + \cdots + N_k
\]

(2.4.9)

where

\[
N_i = \max(n_0+1, \lfloor s_i^2/z \rfloor + 1),
\]

(2.4.10)

hence the expected total sample size is

\[
E(N_{T_1}) = \sum_{i=1}^{k} E(N_i).
\]

(2.4.11)

It follows from Stein (1945, page 247) that

\[
(n_0+1) \left[ \frac{\chi^2_{n_0-1}}{\sigma_i^2} - \frac{(n_0-1)z}{\sigma_i^2} \right] + \frac{\sigma_i^2}{z} \left[ \frac{\chi^2_{n_0+1}}{\sigma_i^2} - \frac{(n_0-1)z}{\sigma_i^2} \right] \leq E(N_i)
\]

\[
\leq (n_0+1) \left[ \frac{\chi^2_{n_0-1}}{\sigma_i^2} - \frac{(n_0-1)z}{\sigma_i^2} \right] + \frac{\sigma_i^2}{z} \left[ \frac{\chi^2_{n_0+1}}{\sigma_i^2} - \frac{(n_0-1)z}{\sigma_i^2} \right] + \left[ \frac{\chi^2_{n_0-1}}{\sigma_i^2} - \frac{(n_0-1)z}{\sigma_i^2} \right],
\]

(2.4.12)
which implies that

\[ 0 \leq E(N_{T_1}) = \sum_{i=1}^{k} \left[ (n_i+1) P \left( \chi^2_{n_i-1} < \frac{(n_i^2-1)z}{\sigma_i^2} \right) + \frac{\sigma_i^2}{z} P \left( \chi^2_{n_i+1} > \frac{(n_i^2-1)z}{\sigma_i^2} \right) \right] \]

\[ \leq \sum_{i=1}^{k} P \left[ \chi^2_{n_i-1} > \frac{(n_i^2-1)z}{\sigma_i^2} \right] \]  

(2.4.13)

Lemma (2.4.14). \( \lim_{\max \sigma_i \to 0} E(N_{T_1}) = k(n_0+1) \) and \( \lim_{(\sigma_i \to \infty) \forall i} E(N_{T_1}) = \infty. \)

Proof. If \( \max \sigma_i \to 0 \) then \( \sigma_i \to 0 \) for all \( i \) and

\[ \lim_{\sigma_i \to 0} \left\{ (n_i+1) P \left( \chi^2_{n_i-1} < \frac{(n_i^2-1)z}{\sigma_i^2} \right) + \frac{\sigma_i^2}{z} P \left( \chi^2_{n_i+1} > \frac{(n_i^2-1)z}{\sigma_i^2} \right) \right\} = n_0+1 \]

since \( \lim_{\sigma_i \to 0} \frac{(n_i^2-1)z}{\sigma_i^2} = \infty. \) Also

\[ \lim_{\sigma_i \to 0} \left\{ E(N_i) - \left\{ (n_i+1) P \left( \chi^2_{n_i-1} < \frac{(n_i^2-1)z}{\sigma_i^2} \right) \right. \right. \]

\[ + \left. \left. \frac{\sigma_i^2}{z} P \left( \chi^2_{n_i+1} > \frac{(n_i^2-1)z}{\sigma_i^2} \right) \right\} \right\} = 0 \]

hence \( \lim_{\sigma_i \to 0} E(N_i) = n_0+1 \) and therefore \( \lim_{\max \sigma_i \to 0} E(N_{T_1}) = k(n_0+1). \)

Letting \( \sigma_i \to \infty \) for any \( i \), the lower bound on \( E(N_i) \) given by (2.4.12) converges to infinity, hence \( \lim_{(\sigma_i \to \infty) \forall i} E(N_{T_1}) = \infty. \)

Lemma (2.4.15). \( E(N_{T_1}) \) is monotonically increasing in each \( \sigma_i^2 \).

Proof. From Seelbinder (1953, p. 642) \( E(N_i) \) is monotonically increasing in \( \sigma_i^2 \) and the result follows since \( E(N_{T_1}) = \sum_{i=1}^{k} E(N_i). \)
Definition (2.4.16). The family of cumulative distribution functions $F_\theta(x)$, indexed by the parameter $\theta$, is said to be stochastically increasing in $\theta$ if for all $\theta, \theta', \theta < \theta'$, $F_{\theta'}(x) \leq F_\theta(x)$ for all $x$. If the random variable $X \sim F_\theta(x)$ then $X$ is said to be stochastically increasing in $\theta$.

Lemma (2.4.17). Let $X \sim F_\theta(x)$ be stochastically increasing in $\theta$ and let $Y \sim G_\eta(y)$ be independent of $X$ and stochastically increasing in $\eta$. Then $X + Y$ is stochastically increasing in $\theta$ and $\eta$.

Proof: Let $\theta < \theta'$ and let $g(y)$ be the density function for $Y$

$$P_{\theta'}[X + Y \leq z] = \int P_{\theta'}[X \leq z - y] g(y) \, dy$$

$$\leq \int P_\theta[X \leq z - y] g(y) \, dy$$

$$= P_\theta[X + Y \leq z]$$

The argument is symmetric in $X$ and $Y$, hence the result follows. (Note that this lemma extends to any finite number of random variables.)

Lemma (2.4.18). The total sample size $N_{T,1}$ is stochastically increasing in $\sigma_i^2$ ($i = 1, \ldots, k$).

Proof. We note that $P[N_1 \leq m] = P[n_0^2 - 1 < (m(n_0 - 1)z)/\sigma_i^2]$ which is clearly stochastically increasing in $\sigma_i^2$. Since $N_{T,1} = N_1 + \cdots + N_k$ it follows from Lemma (2.4.17) that $N_{T,1}$ is stochastically increasing in each $\sigma_i^2$. 
Choice of the $a_{ij}$

Procedure $P_{B_1}$ requires the choosing of constants $\{a_{ij}\}$
($i = 1, \ldots, k; j = 1, \ldots, N_i$) satisfying the requirements given by
(2.1.5a)-(2.1.5c). We suggest choosing the constants $\{a_{ij}\}$ according
to the method given by Chapman and Dudewicz, Ramberg and Chen. This
procedure is as follows: for $i = 1, \ldots, k$ in addition to requiring
$a_{11} = \cdots = a_{iN_0} = a_i$ also set $a_{iN_0+1} = \cdots = a_iN_i = b_i$, then

$$b_i = \frac{1}{N_i} \left( 1 + \sqrt{\frac{n_0(N_i z - s_i^2)}{(N_i - n_0) s_i^2}} \right)$$

(2.4.19)

and

$$a_i = \frac{1 - (N_i - n_0)b_i}{n_0}$$

(2.4.20)

(Note, Dudewicz, Ramberg and Chen give robustness reasons for their
choice.)

The topics of expected sample size and choice of the constants
$\{a_{ij}\}$ were considered only for the case of the one-way layout to avoid
notational difficulties. The generalizations of these topics to the
$r$-way layout follow directly by considering the properties of the expected
total sample size and choice of the constants for each cell of the $r$-way
layout individually.
2.4.A. Multiple-Comparisons

The general theory for multiple-comparisons given in Section 2.2. leads to the following corollaries for the one-way layout as defined by (2.4.1).

Corollary (2.4.21). The probability is $1-\alpha$ that the $k(k-1)/2$ differences $\{\mu_i - \mu_j\}$ are simultaneously covered by the family of intervals

$$\left\{ \tilde{X}_i - \tilde{X}_j \pm \sqrt{z} q_{k, n_0-1}^{\alpha} \right\}$$

where $q_{k, n_0-1}^{\alpha}$ is the upper $\alpha$th percent point of the range of $k$ i.i.d. Student’s-t variates with $n_0-1$ degrees of freedom.

Corollary (2.4.22). The probability $1-\alpha$ that all contrasts $\sum_{i=1}^{k} c_i \mu_i$ are simultaneously covered by the family of intervals

$$\left\{ \sum_{i=1}^{k} c_i \tilde{X}_i \pm \sqrt{z} q_{k, n_0-1}^{\alpha} \frac{1}{2} \sum_{i=1}^{k} |c_i| \text{ for all } \mathbf{c} \in \mathbb{R}^k \right\}$$

Corollary (2.4.23). The probability is $1-\alpha$ that all linear combinations $\sum_{i=1}^{k} \ell_i \mu_i$ are simultaneously covered by the family of intervals

$$\left\{ \sum_{i=1}^{k} \ell_i \tilde{X}_i \pm \sqrt{z} q_{k, n_0}^{\alpha} M(\ell_1, \ldots, \ell_k) \text{ for all } \ell \in \mathbb{R}^k \right\}$$

where $q_{k, n_0}^{\alpha}$ is the upper $\alpha$th percent point of the augmented range of $k$ i.i.d. Student-t variates with $n_0-1$ degrees of freedom.
In practice one may replace $q_{k,n_0}$ with $q_{k,n_0-1}$ which is a good approximation. Tables for $q_{k,n_0-1}$ may be found in Hochberg (1975). The Scheffé-type multiple-comparisons derived from Theorems (2.2.9) and (2.2.10) for the one-way layout are expressed in the following corollaries.

**Corollary (2.4.24).** The probability is $1-\alpha$ that all linear combinations \[ \sum_{i=1}^{k} \ell_i \mu_i \] are simultaneously covered by the family of intervals

\[
\left\{ \sum_{i=1}^{k} \ell_i \tilde{X}_i \pm \sqrt{z_{k,n_0-1}^{\frac{1}{2}} \left( \sum_{i=1}^{k} \ell_i^2 \right)^{1/2}} \right\} \text{ for all } \ell \in \mathbb{R}^k,
\]

where $z_{k,n_0-1}^{\alpha}$ is the upper $\alpha$th percent point for the distribution of \[ \sum_{i=1}^{k} \tilde{t}_i \] where \{\tilde{t}_i\}_{i=1}^{k} are i.i.d. Student's-t variates with $n_0-1$ degrees of freedom.

**Corollary (2.4.25).** The probability is at least $1-\alpha$ that all contrasts \[ \sum_{i=1}^{k} c_i \mu_i \] are simultaneously covered by the family of intervals

\[
\left\{ \sum_{i=1}^{k} c_i \tilde{X}_i \pm \sqrt{z_{k,n_0-1}^{\frac{1}{2}} \left( \sum_{i=1}^{k} c_i^2 \right)^{1/2}} \right\} \text{ for all } \ell \in \mathscr{L}_C
\]

where $z_{k,n_0-1}^{\alpha}$ is the upper $\alpha$th percent point of the distribution of $Q = \sum_{i=1}^{k} (t_i - \bar{t})^2$, \{\tilde{t}_i\}_{i=1}^{k} are i.i.d. Student's-t variates with $n_0-1$ degrees of freedom, and $\bar{t} = 1/k \sum_{i=1}^{k} t_i$.

In the case of the one-way layout (but not for higher-way layouts) it should be noted that Hochberg (1975) provided simultaneous intervals for all pairwise comparisons, contrasts and linear combinations
His procedures correspond to our corollaries (2.4.23), (2.4.24), and (2.4.25). In these cases he used the overall sample mean \( \bar{x}_1 \) rather than \( \tilde{x}_1 \), and his intervals are of random lengths but with probability one are less than our lengths. However, Hochberg's procedure does not generalize to higher way layouts, nor does his procedure generate theorems analogous to our corollaries (2.4.21) and (2.4.22).

The sample sizes required by our procedure and Hochberg's respectively for the \( i \)th population are

\[
N_i = \max\{n_0 + 1, \lfloor s_i^2 / z \rfloor + 1\}
\]

and

\[
N_i^H = \max\{n_0, \lfloor s_i^2 / z \rfloor + 1\} \quad (i = 1, \ldots, k)
\]

Thus,

\[
N_i \leq N_i^H + 1
\]

and

\[
\mathbb{E}(N_i) \leq \mathbb{E}(N_i^H) + 1 \quad (i = 1, \ldots, k)
\]

The exact relationship depends upon the unknown \( \sigma_i^2 \) but as we have previously discussed \( \lim_{\sigma_i^2 \to 0} \mathbb{E}(N_i) = n_0 + 1 \) and similarly

\[
\lim_{\sigma_i^2 \to 0} \mathbb{E}(N_i^H) = n_0.
\]

Thus for small \( \sigma_i^2 \), \( \mathbb{E}(N_i^H) < \mathbb{E}(N_i) \), while letting \( \sigma_i^2 \to \infty \) we find \( \mathbb{E}(N_i) = \mathbb{E}(N_i^H) \).
2.5 Procedure $\mathcal{P}_{B_2}$

We next consider a procedure for testing (2.4.2) which is motivated by the work of Chatterjee (1959a). A critical analysis of that work shows that Chatterjee is essentially developing estimators of the elements of the mean vector of the multivariate normal distribution $N_p(y, \Sigma)$ which when conditioned on the first-stage sample estimators of the variances and covariances are independent with the same variance. The conditioning variables are then replaced by an equivalent set of variables which are independent of $\Sigma$, which leads to the test statistic (1.3.19) which has a distribution independent of $\Sigma$. Procedure $\mathcal{P}_{B_2}$ modifies this idea and applies it to testing (2.4.2).

Let $\pi_i$ refer to the population defined by the $i$th factor in the one-way layout. Procedure $\mathcal{P}_{B_2}$ works as follows: apply the two-stage sampling scheme $\mathcal{R}(n_0, 2k; s_i^2, 2)$ to $\pi_i$ where $s_i^2$ is the usual unbiased estimator of $\sigma_i^2$ based on the first $n_0$ observations. For each $r = 1, \ldots, k$ select constants $\{a_{rij}\}$ ($i = 1, \ldots, k, j = 1, \ldots, N_i$) such that

$$a_{ril} = \cdots = a_{rin_0},$$

$$(2.5.1a)$$

$$\sum_{j=1}^{N_i} a_{rij} = \begin{cases} 0, & i \neq r \\ 1, & i = r, \end{cases}$$

$$(2.5.1b)$$

and
where \( N_i \) is the final sample size for \( \pi_i \). Define

\[
\xi_r = \sum_{i=1}^{k} \sum_{j=1}^{N_i} a_{rij} x_{ij}, \quad r = 1, \ldots, k. \tag{2.5.2}
\]

Given \( s_1^2, \ldots, s_k^2, N_i \) and \( a_{rij} \) are fixed, so

\[
E(\xi_r) = \sum_{i=1}^{k} \sum_{j=1}^{N_i} a_{rij} \mu_i = \sum_{i=1}^{k} (\mu_i \sum_{j=1}^{N_i} a_{rij}) = \mu_r, \tag{2.5.3}
\]

\[
\text{Var}(\xi_r) = \sum_{i=1}^{k} \sum_{j=1}^{N_i} a_{rij}^2 \sigma_i^2 = \sum_{i=1}^{k} \sigma_i^2 (\sum_{j=1}^{N_i} a_{rij}^2) = \sum_{i=1}^{k} \sigma_i^2 \frac{a_{rij}^2}{s_i^2} = zL \tag{2.5.4}
\]

where \( L = \sum_{i=1}^{k} \sigma_i^2 / s_i^2 \) and

\[
\text{Cov}(\xi_r, \xi_s) = \sum_{i=1}^{k} \sum_{j=1}^{N_i} a_{rij} a_{sij} \sigma_i^2 = 0 \quad (r \neq s). \tag{2.5.5}
\]

It now follows that \( \xi_1, \ldots, \xi_k \) are conditionally (given \( s^2 = (s_1^2, \ldots, s_k^2) \)) independent normal random variables with a common variance given by (2.5.4). Let \( B = (b_{ij}) \) be defined by (2.4.4) and define
\begin{align}
F' &= \sum_{i=1}^{k} \sum_{j=1}^{k} \frac{b_{ij} i^i j^j}{z}.
\end{align}

\text{(2.5.6)}

Conditionally, by Lemma (2.1.17) \( F'/L \) is distributed as a \( \chi^2_{k-1}(\Delta^*) \) with
\[ \Delta^* = \frac{\Delta}{zL} \]
where
\[ \Delta = \sum_{i=1}^{k} (\mu_i - \bar{\mu})^2. \]
\text{(2.5.7)}
\text{(2.5.8)}

The conditional density for \( F'/L \), from (1.2.1), is given by
\begin{align}
f(x|s) &= \frac{e^{-\left(\frac{\Delta}{2zL}\right)}}{2^{(k-1)/2}} \sum_{j=0}^{\infty} \frac{\left(\frac{\Delta}{zL}\right)^j}{\Gamma\left(\frac{k-1}{2} + j\right)2^j j!} \frac{x^{(k-1)/2+j-1} e^{-x/2}}{\Gamma\left(\frac{k-1}{2} + j\right) 2^j j!} 
\end{align}
\text{(0 < x < \infty)}
\text{(2.5.9)}

hence the conditional density of \( F' \) is, for \( 0 < x < \infty \), given by
\begin{align}
g(x|s) &= \frac{e^{-\left(\frac{\Delta}{2zL}\right)}}{L 2^{(k-1)/2}} \sum_{j=0}^{\infty} \frac{\left(\frac{\Delta}{zL}\right)^j}{\Gamma\left(\frac{k-1}{2} + j\right) 2^j j!} \frac{x^{(k-1)/2+j-1} e^{-x/2L}}{\Gamma\left(\frac{k-1}{2} + j\right) 2^j j!} 
\end{align}
\text{(2.5.10)}

Noting that \( V_i = \frac{(n_0-1)s_i^2}{\sigma_i^2} \sim \chi^2_{n_0-1} \), we replace \( L \) by \( (n_0-1) \sum_{i=1}^{k} 1/V_i \) and have
The conditional density of $F'$ involves the $s_i^2$ only through
\[ \sum_{i=1}^{k} \frac{1}{V_i}, \text{ hence only through the } \{V_i\}_{i=1}^{k}, \text{ which are i.i.d. } \chi_{n_0-1}^2 \text{ variates.} \]
The joint density of $\{V_i\}_{i=1}^{k}$ is given by
\[
\begin{equation}
\begin{aligned}
\{ \begin{array}{ll}
\frac{1}{ \prod_{i=1}^{k} \Gamma(\frac{n_0-1}{2}) } & \frac{1}{(n_0-1)/2-1} \frac{1}{v_i} e^{-v_i/2} \\
0 & \text{otherwise} \\
\end{array} \right.
\end{aligned}
\end{equation}
\]
The unconditional density of $F'$ may be obtained by integrating (2.5.11) with respect to (2.5.12). Thus the unconditional density of $F'$ may be expressed as
\[
\begin{equation}
g(x) = \int_{0}^{\infty} \cdots \int_{0}^{\infty} g(x|s) \prod_{i=1}^{k} \frac{1}{\Gamma(\frac{n_0-1}{2})} \frac{1}{(n_0-1)/2-v_i} e^{-v_i/2} dv_1 \cdots dv_k.
\end{equation}
\]
Now (2.5.13) is the general ("noncentral") unconditional distribution of $F'$. Under the null hypothesis $H_0 : \mu_1 = \cdots = \mu_k, \Delta = 0$ and (2.5.13) reduces to the "central" distribution of $F'$ given by
\[
g(x) = \frac{1}{2^{(k-1)/2} \Gamma\left(\frac{k-1}{2}\right)(n_0-1)(k-1)/2} 
\cdot \int_0^\infty \cdots \int_0^\infty x^{(k-3)/2} \exp\left(-\frac{x}{2(n_0-1)} \sum_{i=1}^k \frac{1}{\nu_i}\right) \cdot \left(\sum_{i=1}^k \frac{1}{\nu_i}\right)^{-(k-1)/2} 
\cdot \prod_{i=1}^k \frac{1}{\Gamma\left(\frac{n_0-1}{2}\right)\nu_i^{(n_0-3)/2}} \exp\left(-\frac{x}{\nu_i}\right)^{-\frac{n_0-3}{2}} 
\cdot e^{-v/2} dv_1 \cdots dv_k \quad (2.5.14)
\]

We take as our test statistic \( F' \) and reject (2.4.2) if and only if \( F' > F'_{k-1, \alpha} \) where \( F'_{k-1, \alpha} \) is the upper \( \alpha \)-th percent point of the central distribution of \( F' \) given by (2.5.14).

Existence of \( \{a_{rij}\} \)

We demonstrate the existence of constants satisfying (2.5.1a)-(2.5.1c) by displaying explicitly such a set of constants. Define

\[
a_{rij} = \begin{cases} 
0, & i \neq r, 1 \leq j \leq n_0, j \neq 2r-1 + n_0, j \neq 2r + n_0 \\
\sqrt{z/s_i}, & i \neq r, j = 2r-1 + n_0 \\
-\sqrt{z/s_i}, & i \neq r, j = 2r + n_0 \\
a_i, & i = r, 1 \leq j \leq n_0 \\
b_i, & i = r, n_0+1 \leq j \leq N_i,
\end{cases} \quad (2.5.15)
\]

where \( a_i \) and \( b_i \) are chosen using (2.4.19) and (2.4.20) (note the discussion of robustness there applies here also). Then (2.5.1a)-(2.5.1c) are satisfied.
Properties of $P_{B_2}$

As with $P_{B_1}$, we require $P_{B_2}$ to possess certain properties if it is to be considered a solution to our problem. In particular, the distribution of $F'$ must be independent of the unknown variances and the power of the test based on $F'$ must be controllable.

Lemma (2.5.16). The distribution of $F'$ under the null or alternative hypotheses is independent of $\sigma_1^2, \ldots, \sigma_k^2$.

Proof. From (2.5.13) it follows that the distribution of $F'$ depends upon the parameter $\Delta = \sum_{i=1}^{k} (\mu_i - \bar{u})^2$ only and hence is independent of $\sigma_1^2, \ldots, \sigma_k^2$.

The power function of the test based on $F'$ is a function of the means only through $\Delta = \sum_{i=1}^{k} (\mu_i - \bar{u})^2$, hence we denote it by

$$\beta(\Delta) = P[F' > F'_{k-1, \alpha}/\Delta]. \quad (2.5.17)$$

We next prove that the power function has the important property of being monotonically increasing in $\Delta$.

Lemma (2.5.18). The power function $\beta(\Delta)$ is monotonically increasing in $\Delta$.

Proof. $\beta(\Delta) = \int P[F' > F'_{k-1, \alpha}/s] f(s) \, ds$

where $f(s)$ is the joint density of $s = (s_1^2, \ldots, s_k^2)$ and

$P[F' > F'_{k-1, \alpha}/s]$ denotes the conditional power given $s = (s_1^2, \ldots, s_k^2)$.
and the integral is taken over the range of the $s_i^2$'s. Thus

$$\beta(\Delta) = \int P \left[ \frac{F'}{L} > \frac{F'_{k-1, \alpha/L}}{s} \right] f(s) \, ds$$

and conditionally $F'/L$ is distributed as $\chi^2_{k-1}(\Delta/zL)$. It is well-known that $P[F'/L > (F'_{k-1, \alpha/L})/s]$ is increasing in $\Delta/zL$, hence in $\Delta$, and therefore $\beta(\Delta)$ is increasing in $\Delta$.

**Corollary (2.5.19).** The test of $(2.4.2)$ based on $F'$ is unbiased.

**Corollary (2.5.20).** $\lim_{z \to \infty} \beta(\Delta) = 1$, hence the power function $\beta(\Delta)$ is controllable through $z$.

**Expected Sample Size**

Let $N_{T_2}$ denote the total sample size for $\chi^2_{B_2}$,

$$N_{T_2} = N_1' + \cdots + N_k'$$

where

$$N_1' = \max\{n_0+2k, \lceil s_1^2/z \rceil + 1\}.$$

$N_1'$ may assume any of the values $\{n_0+2k, n_0+2k+1, \ldots\}$ and

$$P[N_1' = n_0 + 2k] = P[s_1^2/z \leq n_0 + 2k]$$

$$= P \left[ \frac{(n_0-1)s_1^2}{\sigma_1^2} \leq \frac{(n_0+2k)(n_0-1)z}{\sigma_1^2} \right]$$

$$= P \left[ \chi^2_{n_0-1} \leq \frac{(n_0+2k)(n_0-1)z}{\sigma_1^2} \right]. \quad (2.5.21)$$
For any integer $m > n_0 + 2k$,

$$P[N'_i = m] = P[m-1 < s_i^2/z < m]$$

$$= P \left[ \frac{z(m-1)(n_0-1)}{\sigma_i^2} < \chi^2_{n_0-1} \leq \frac{zm(n_0-1)}{\sigma_i^2} \right]. \quad (2.5.22)$$

Let $f(u)$ denote the density for a $\chi^2_{n_0-1}$ random variable. Then

$$P[N'_i = n_0 + 2k] = \int_0^{(n_0+2k)(n_0-1)z/\sigma_i^2} f(u) \, du \quad (2.5.23)$$

and

$$P[N'_i = m] = \int_{z(m-1)(n_0-1)/\sigma_i^2}^{zm(n_0-1)/\sigma_i^2} f(u) \, du. \quad (2.5.24)$$

Further, over the interval $(z(m-1)(n_0-1)/\sigma_i^2, zm(n_0-1)/\sigma_i^2)$,

$$\frac{z(m-1)(n_0-1)}{\sigma_i^2} < \mu < \frac{zm(n_0-1)}{\sigma_i^2},$$

hence

$$\frac{\sigma_i^2u}{z(n_0-1)} < m < \frac{\sigma_i^2u}{z(n_0-1)} + 1.$$ 

Now

$$E(N'_i) = \sum_{m=n_0+2k}^{\infty} mP[N'_i = m], \quad (2.5.25)$$

so
\[
E(N'_1) = \int_0^{\infty} \frac{z(n_0+2k)(n_0-1)/\sigma_i^2}{z(n_0+2k)(n_0-1)/\sigma_i^2} \left( n_0+2k \right) f(u) \, du \\
\quad + \sum_{m=n_0+2k+1}^{\infty} \int_{m(n_0-1)/\sigma_i^2}^{\infty} \frac{z(n_0-1)/\sigma_i^2}{z(m-1)(n_0-1)/\sigma_i^2} \left( m-1 \right) f(u) \, du
\]

\[
z(n_0+2k)(n_0-1)/\sigma_i^2 \leq \int_0^{\infty} \frac{z(n_0+2k)(n_0-1)/\sigma_i^2}{z(n_0+2k)(n_0-1)/\sigma_i^2} \left( n_0+2k \right) f(u) \, du \\
\quad + \int_0^{\infty} \frac{\sigma_i^2 u}{z(n_0-1)} \, f(u) \, du.
\]

Similarly,

\[
z(n_0+2k)(n_0-1)/\sigma_i^2 \leq \int_0^{\infty} \frac{z(n_0+2k)(n_0-1)/\sigma_i^2}{z(n_0+2k)(n_0-1)/\sigma_i^2} \left( n_0+2k \right) f(u) \, du \\
\quad + \int_0^{\infty} \frac{\sigma_i^2 u}{z(n_0-1)} \, f(u) \, du.
\]

Summing chi-square integrals as in Stein (1945, p. 247), it follows that

\[
0 \leq E(N'_1) - (n_0+2k) P \left[ \chi^2_{n_0-1} < \frac{z(n_0+2k)(n_0-1)}{\sigma_i^2} \right] - \frac{\sigma_i^2}{z} P \left[ \chi^2_{n_0+1} > \frac{z(n_0+2k)(n_0-1)}{\sigma_i^2} \right]
\]

\[
\leq P \left[ \chi^2_{n_0-1} > \frac{z(n_0+2k)(n_0-1)}{\sigma_i^2} \right].
\]

Since \( E(N'_{T_2}) = \sum_{i=1}^{k} E(N'_1) \),
\[ 0 \leq E(N_{T_2}) - \sum_{i=1}^{k} \left( \frac{n_0+2k}{\sigma_i^2} \right) \left[ \frac{\chi^2_{(n_0+2k)(n_0-1)}}{\chi^2_{(n_0-1)}} \right] \]

\[ + \frac{\sigma_i^2}{z} \cdot P \left[ \frac{\chi^2_{(n_0+2k)(n_0-1)}}{\sigma_i^2} \right] \]

\[ \leq \sum_{i=1}^{k} P \left[ \frac{\chi^2_{(n_0+2k)(n_0-1)}}{\sigma_i^2} \right] \] (2.5.29)

**Lemma (2.5.30).** \( \lim_{\sigma_i \to 0} E(N_{T_2}) = k(n_0+2k) \) and \( \lim_{\sigma_i \to \infty} \) for any \( i \) \( E(N_{T_2}) = \infty. \)

**Proof.** Analogous to the proof for Lemma (2.4.14).

**Lemma (2.5.31).** \( E(N_{T_2}) \) is monotonically increasing in each \( \sigma_i^2 \).

**Proof.** Analogous to the proof for Lemma (2.4.15).

**Lemma (2.5.32).** \( N_{T_2} \) is stochastically increasing in each \( \sigma_i^2 \).

**Proof.** Analogous to the proof of Lemma (2.4.18).

**Limiting Distribution of \( F' \).**

The exact distribution of \( F' \) is not yet tabulated and an approximation is needed. We consider approximating the exact distribution of \( F' \) by its limiting distribution as \( n_0 \to \infty. \)
Theorem (2.5.33). The limiting distribution of $F'/k$ as $n_0 \to \infty$ is

$$
\chi^2_{k-1}(\Delta) \quad \text{where} \quad \Delta = \sum_{i=1}^{k} \left( \mu_i - \bar{\mu} \right)^2 / kz.
$$

Proof. The conditional joint distribution of $\xi_1, \ldots, \xi_k$ given $s^2 = (s_1^2, \ldots, s_k^2)$ is $N_k(\mu, (z \sum_{i=1}^{k} \sigma_i^2 / s_i^2)I)$ where $\mu = (\mu_1, \ldots, \mu_k)'$.

For $i = 1, \ldots, k$, $s_i^2$ converges in probability to $c_i^2$ and so $\sum_{i=1}^{k} c_i^2 / s_i^2$ converges in probability to $k$ as $n_0 \to \infty$. Therefore by the Helly-Bray Theorem (see page 84 of Tucker (1967)) the limiting joint distribution of $\xi_1, \ldots, \xi_k$ as $n_0 \to \infty$ is $N_k(\mu, zkI)$ and thus the limiting joint distribution of $(\xi_1 / \sqrt{zk}, \ldots, \xi_k / \sqrt{zk})$ is $N_k((1/\sqrt{zk})\mu, I)$.

Now $F'/k$ is a continuous function of $(\xi_1 / \sqrt{zk}, \ldots, \xi_k / \sqrt{zk})$ and so by (2.1.16) $F'/k$ is distributed as $X'BX$ as $n_0 \to \infty$ where $X \sim N_k((1/\sqrt{zk})\mu, I)$ and $B$ is defined by (2.4.4). It then follows by Lemma (2.1.17) that $F'/k$ is distributed as $\chi^2_{k-1}(\Delta)$ with $\Delta = \sum_{i=1}^{k} \left( \mu_i - \bar{\mu} \right)^2 / kz$. 

2.6. Procedure \( \mathcal{P}_{B_3} \)

In this section we obtain a third method of testing (2.4.2), again with the property that the test is completely free of the unknown variances. Note that Chakravarti (1965) does provide a test of (2.4.2) such that the level of the test is independent of the unknown variances, however the power is not. His procedure generalizes the work of Scheffe (1943) and forces the problem of testing (2.4.2) into an equivalent problem of testing a \((k-1)\) dimensional mean vector is equal to \( \mathbb{Q} \). Chatterjee (1959a) extended Stein's work to test just such hypotheses (mean vectors equal to \( \mathbb{Q} \)) such that the power was independent of the unknown variances and covariances. We propose to combine the two procedures in such a way so as to yield a test of (2.4.2) which is completely independent of the unknown variances.

Initially choose \( n_0 > k-1 \) observations from each population and let \((X_{11}, \ldots, X_{1n_0})\) be the observations from \( \pi_1 \). Next compute the Scheffe' variables (see Scheffe' (1943))

\[
Y_{j\alpha} = X_{1\alpha} - X_{j\alpha} \quad (\alpha = 1, \ldots, n_0; \ j = 2, 3, \ldots, k), \tag{2.6.1}
\]

then

\[
\mathbb{E}[Y_{j\alpha}] = u_1 - u_j, \quad \text{Var}(Y_{j\alpha}) = \sigma_1^2 + \sigma_i^2, \tag{2.6.2}
\]

and

\[
\text{Cov}(Y_{j\alpha}, Y_{j', \alpha'}) = \begin{cases} 
\sigma_1^2 + \sigma_j^2, & \alpha = \alpha', \ j = j' \\
\sigma_1^2, & \alpha = \alpha', \ j \neq j' \\
0, & \text{otherwise}
\end{cases}
\]
The vector $\mathbf{U}_\alpha = (Y_{2\alpha}, \ldots, Y_{k\alpha})'$ is then distributed as a $(k-1)$-dimensional multivariate normal with mean vector

$$E[\mathbf{U}_\alpha] = (\mu_1 - \mu_2, \ldots, \mu_1 - \mu_k)'$$

and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_k^2 \end{pmatrix}$$

with

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_k^2 \end{pmatrix}$$

for $i \neq j$

$$\sigma_{ij} = \begin{cases} \sigma_i^2, & i \neq j \\ \sigma_i^2 + \sigma_j^2, & i = j \end{cases}$$

for $\alpha = 1, 2, \ldots, n_0$.

$\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_{n_0}$ may then be considered as a random sample of size $n_0$ from this multivariate distribution. The hypothesis of interest, $H_0 : \mu_1 = \cdots = \mu_k$, is equivalent to the hypothesis that the mean vector of the above distribution is $0$. Applying Chatterjee's results, we compute

$$\bar{\mathbf{y}}_j = \frac{1}{n_0} \sum_{\alpha=1}^{n_0} Y_{j\alpha}$$

and

$$s_{ij} = \frac{1}{n_0} \sum_{\alpha=1}^{n_0} (Y_{i\alpha} - \bar{\mathbf{y}}_i)(Y_{j\alpha} - \bar{\mathbf{y}}_j)$$

and

$$s_{ij}' = \frac{1}{n_0-1} s_{ij}$$

for $i, j = 2, 3, \ldots, k$.

Next apply $D(n_0, (k-1)^2; \sum_{i=2}^{k} \sum_{j=2}^{k} \alpha_{ij}s_{ij}; z)$ to this multivariate normal population, where $(\alpha_{rs})$ is a $(k-1) \times (k-1)$ positive definite matrix. Take $N-n_0$ additional observations on each population and compute the Scheffé variables with these additional observations:
\[ Y_{j\alpha} = X_{j\alpha} - X_{j\alpha}, \quad \alpha = n_0 + 1, \ldots, N \quad (2.6.9) \]

Let \( k' = k - 1 \) and choose \( k' \ (k' \times N) \) matrices

\[
A_r = \begin{bmatrix}
        a_{r11} & \cdots & a_{r1N} \\
        \vdots & & \vdots \\
        a_{rk'1} & a_{rk'N}
\end{bmatrix}, \quad r = 2, \ldots, k \quad (2.6.10)
\]

(note as previously discussed Chatterjee showed such matrices exist by our choice of \( N \)), such that

a) \( a_{ril} = \cdots = a_{rin_0} \), \quad r, i = 2, 3, \ldots, k, \quad (2.6.11a)

b) \[
\sum_{m=1}^{N} a_{rim} = \begin{cases} 
1, & i = r, \\
0, & i \neq r,
\end{cases} \quad r, i = 2, 3, \ldots, k \quad (2.6.11b)
\]

c) \[
\sum_{m=1}^{N} a_{rim} a_{sjm} = \delta_{rs} \delta_{ij} \quad (2.6.11c)
\]

The final observation matrix is given by

\[
\begin{bmatrix}
Y_{21} & Y_{22} & \cdots & Y_{2N} \\
Y_{31} & Y_{32} & \cdots & Y_{3N} \\
\vdots & & \vdots \\
Y_{k1} & Y_{k2} & \cdots & Y_{kN}
\end{bmatrix} \quad (2.6.12)
\]
and we construct the $k-1$ variables $\eta_2, \eta_3, \ldots, \eta_k$, where

$$\eta_r = \sum_{i=2}^{k} \sum_{m=1}^{N} \alpha_{rmi} Y_{im}, \quad r = 2, \ldots, k, \quad (2.6.13)$$

then take as test statistic

$$F'' = \frac{1}{2} \sum_{r=2}^{k} \sum_{s=2}^{k} \alpha_{rs} \eta_r \eta_s. \quad (2.6.14)$$

From the results of Chatterjee (1959a), page 127 the distribution of $F''$ is

$$f(u) = \frac{1}{n_0^{-1}} \sum_{t=0}^{\infty} \frac{1}{t!} \left\{ \left( \frac{\Delta}{2(n_0-1)} \right)^t \right\} 2^{k'-2t} \frac{1}{\Gamma \left( \frac{k'+2t}{2} \right)} \left\{ \frac{u}{n_0-1} \right\}^{(k'+2t)/2-1} \times \int_{0<\lambda_1<\cdots<\lambda_k<\infty} \exp \left\{ -\frac{u + \Delta}{2(n_0-1)} \left( \sum_{i=1}^{k'} \frac{1}{\lambda_i} \right) \right\} \left( \sum_{i=1}^{k'} \frac{1}{\lambda_i} \right)^{-\left( k'+4t \right)/2} \exp(\lambda_1, \ldots, \lambda_k) \, d\lambda_1 \cdots d\lambda_k, \quad 0 < \lambda_k, < \cdots < \lambda_1 < \infty \quad (2.6.15)$$

where $\lambda_1, \ldots, \lambda_k$ are the characteristic roots of the matrix $(\sigma_{ij})(s_{ij}^{ij})$, and their joint density, derived by Roy (1939) and Hsu (1939), is given by

$$p(\lambda_1, \ldots, \lambda_k) = 2^{-(k'(n_0-1))/2} \prod_{i=1}^{k'} \frac{\Delta_1}{\Gamma \left( \frac{n_0-k'}{2} \right)} \Gamma \left( \frac{k'-i+1}{2} \right) \left( \frac{n_0-k'-2}{2} \right)^{(n_0-k'-2)/2} \exp \left\{ -\sum_{i=1}^{k'} \lambda_i \right\} \left\{ \prod_{i=1}^{k'} \lambda_i \right\}^{(n_0-k'-2)/2} \left\{ \prod_{i=1}^{k'} \prod_{j=i+1}^{k'} (\lambda_i - \lambda_j) \right\} 0 < \lambda_k, < \cdots < \lambda_1 < \infty. \quad (2.6.16)$$
The noncentrality parameter of the distribution is

$$\Delta = \frac{1}{2} \sum_{r=2}^{k} \sum_{s=2}^{k} \alpha_{rs}(\mu_{1r} - \mu_{1s})(\mu_{1r} - \mu_{1s}).$$  \hspace{1cm} (2.6.17)$$

Under \( H_0 : (\mu_{1r}, \ldots, \mu_{1k}) = 0 \), \( \Delta = 0 \); but, otherwise \( \Delta > 0 \).

The choice of \( (\alpha_{rs}) \) is arbitrary but may be used to the advantage of the experimenter. For example if it is more important to detect a difference between \( \mu_{ir} \), and \( \mu_{il} \) or between \( \mu_{is} \), and \( \mu_{il} \) we may appropriately choose \( \alpha_{rs} \) to weight \( (\mu_{1r} - \mu_{il}) \) and \( (\mu_{1s} - \mu_{il}) \) in the noncentrality parameter. If no population is more important than the others we suggest taking \( (\alpha_{rs}) \) to be the identity matrix.

We also note that this procedure is advantageous if population \( P_1 \) can be considered as a control. The noncentrality parameter in a certain sense measures how far the other \( k-1 \) populations are from the control.

Properties of \( \mathcal{P}_{B_3} \)

As with procedures \( \mathcal{P}_{B_1} \) and \( \mathcal{P}_{B_2} \), we will consider \( \mathcal{P}_{B_3} \) a solution to our problem if it possesses certain properties (i.e. it is independent of the unknown variances).

**Lemma (2.6.18).** The distribution of \( F'' \) is independent of the unknown variances under both null and alternative hypotheses.
Proof. The distribution of $F''$ depends only upon

$$\Delta = 1/z \sum_{r=2}^{k} \sum_{s=2}^{k} \alpha_{rs}(\mu_{1-r})(\mu_{1-s})$$

which is independent of the unknown variances.

Lemma (2.6.19). The power function of the test of (2.4.2) based on $F''$ is a strictly increasing function of $\Delta$.

Proof. Conditionally on $(s_{ij})$

a) $E(\eta_r) = \mu_{1-r}$

b) $V(\eta_r) = z\alpha_{rr} \sum_{i=2}^{k} \sum_{j=2}^{k} \sigma_{ij}^{s_{ij}}$

and

c) $\text{cov}(\eta_r, \eta_s) = z\alpha_{rs} \sum_{i=2}^{k} \sum_{j=2}^{k} \sigma_{ij}^{s_{ij}}$

Letting $L = \sum_{i=2}^{k} \sum_{j=2}^{k} \sigma_{ij}^{s_{ij}}$, the conditional covariance matrix of $\eta_2, \ldots, \eta_k$ is $zL(\alpha_{rs})$, hence by Lemma (2.1.17) $F''/L$ conditionally is distributed as $\chi_{k-1}^2(\Delta/L)$. Let $\beta(\Delta)$ denote the power of the test based on $F''$. Then

$$\beta(\Delta) = \int P[F'' > \frac{F''}{\alpha_{s_{ij}, \Delta}}, p(s)ds$$

$$= \int P[F'' > \frac{\alpha_{s_{ij}, \Delta}}{L}, p(s)ds,$$

where $F''$ is the upper $\alpha$s percent point of the unconditional central $F''$ distribution and $p(s)$ is the joint density of the $s_{ij}$. Let $\Delta_1 < \Delta_2$. Then, uniformly in the $s_{ij}$,
\[ P\left[ \frac{F''}{L} > \frac{F''}{\alpha L} \mid s_{ij}, \Delta \right] \leq P\left[ \frac{F''}{L} > \frac{F''}{\alpha L} \mid s_{ij}, \Delta_2 \right] \]

since the noncentral chi-square distribution is monotone increasing in \( \Delta \). Since this holds uniformly in the \( s_{ij} \)

\[ \beta(\Delta) = \int P\left[ F''/L, F''/\alpha L \mid s_{ij}, \Delta_1 \right] p(s) \, ds \]

\[ \leq \int P\left[ F''/L > F''/\alpha L \mid s_{ij}, \Delta_2 \right] p(s) \, ds = \beta(\Delta_2). \]

From this Lemma (2.6.19) we see that the power is controllable through \( z \) since it follows from Lemma (2.6.19) that

\[ \lim_{z \to \infty} \beta(\Delta) = 1, \quad (2.6.20) \]

and this limit is monotone in \( z \).

**Expected Sample Size**

The total sample size for procedure \( \mathcal{B}_3 \) is \( N = kN \) where

\[ N = \max\{n_0 + (k-1)^2, \sum_{i=2}^{k} \sum_{j=2}^{k} \alpha_{ij} s_{ij} \} + 1, \]

hence the expected total sample size is \( E(N) = kE(N) \). Now the range of \( N \) is \( \{n'_0, n'_0 + 1, \ldots\} \) where \( n'_0 = n_0 + (k-1)^2 \),

\[ P[N = n'_0] = \prod_{z=2}^{k} \sum_{i=2}^{k} \sum_{j=2}^{k} \alpha_{ij} s_{ij} \leq n'_0 \]  \quad (2.6.21)
and (for any integer \( m > n'_0 \))

\[
P[N=m] = P[m-1 < \frac{1}{z} \sum_{i=2}^{k} \sum_{j=2}^{k} s_{ij} \leq m]
\]  

(2.6.22)

We may simplify the preceding procedure if we take the matrix \((\alpha_{rs})\) to be the \((k-1)\times(k-1)\) identity matrix. In this case

\[
N = \max\{n_0 + (k-1)^2, \left[ \frac{1}{z} \sum_{i=2}^{k} s_{ii} \right] + 1 \},
\]  

(2.6.23)

\[
F'' = \frac{1}{z} \sum_{r=2}^{k} \eta_r^2
\]  

(2.6.24)

with

\[
\Delta = \frac{1}{z} \sum_{r=2}^{k} (\mu_r - \mu_1)^2,
\]  

(2.6.25)

we have

\[
P[N=n'_0] = P[\sum_{i=2}^{k} s_{ii} \leq n'_0]
\]  

\[= P[\sum_{i=2}^{k} (n_0-1)(\sigma_1^2 + \sigma_i^2)x_i^2 \leq n'_0]
\]  

(2.6.26)

and (for any integer \( m > n'_0 \))

\[
P[N=m] = P[m-1 < \frac{1}{z} \sum_{i=2}^{k} (n_0-1)(\sigma_1^2 + \sigma_i^2)x_i^2 \leq m],
\]

where \( \{x_i^2\}_{i=2}^{k} \) are \( \chi^2_{n_0-1} \) variates but not independent, since they all involve observations from \( \pi_1 \).

The cumulative distribution function for \( N \) is then given (for any integer \( m \geq n'_0 \)) by
Lemma (2.6.28). The total sample size $N_{T3}$ is stochastically increasing in $\sigma_i^2$ ($i = 1, 2, \ldots, k$).

Proof. From Lemma (2.4.17) and expression (2.6.27) it follows that $N$ is stochastically increasing in $\sigma_i^2$ ($i = 1, 2, \ldots, k$) and, since $N_{T3} = kN$, the result follows.

When $k = 2$,

$$N = \max[n_0 + 1, \left(\frac{1}{z} s_2^2\right) + 1]$$

where

$$s_2^2 = \frac{1}{n_0 - 1} \sum_{\alpha=1}^{n_0} (Y_{2\alpha} - \bar{Y}_2)^2,$$

$$Y_{2\alpha} = X_{1\alpha} - X_{2\alpha},$$

and

$$\bar{Y}_2 = \frac{1}{n_0} \sum_{\alpha=1}^{n_0} Y_{2\alpha},$$

hence

$$\frac{(n_0 - 1)s_2^2}{\sigma_1^2 + \sigma_2^2} \chi^2_{n_0 - 1}, \quad F = \eta_2^2,$$

and

$$\Delta = \frac{(\mu_1 - \mu_2)^2}{z}$$

In this case

$$P[N = n_0 + 1] = P[s_2^2 / z \leq n_0 + 1]$$

$$= P \left[ \chi^2_{n_0 - 1} < \frac{(n_0^2 - 1)z}{\sigma_1^2 + \sigma_2^2} \right]$$

(2.6.32)
and (for $m > n_0 + 1$)

$$P[N = m] = P \left[ \frac{(m-1)(n_0-1)z}{\sigma_1^2 + \sigma_2^2} < \chi^2_{n_0-1} \leq \frac{m(n_0-1)z}{\sigma_1^2 + \sigma_2^2} \right], \quad (2.6.33)$$

hence

$$E(N_{T_3}) = 2(n_0 + 1) P \left[ \frac{\chi^2_{n_0-1}}{\sigma_1^2 + \sigma_2^2} \leq \frac{(n_0-1)z}{\sigma_1^2 + \sigma_2^2} \right]$$

$$+ 2 \sum_{m=n_0+2}^{\infty} m P \left[ \frac{(m-1)(n_0-1)z}{\sigma_1^2 + \sigma_2^2} < \chi^2_{n_0-1} \leq \frac{m(n_0-1)z}{\sigma_1^2 + \sigma_2^2} \right]$$

$$= 2(n_0 + 1) P \left[ \frac{\chi^2_{n_0-1}}{\sigma_1^2 + \sigma_2^2} \leq \frac{(n_0-1)z}{\sigma_1^2 + \sigma_2^2} \right]$$

Bounds on $E(N_{T_3})$ may be generated which are similar to those given in the previous two procedures $P_{B_1}$ and $P_{B_2}$. That is

$$0 \leq E(N_{T_3}) - 2 \left\{ (n_0 + 1) P \left[ \frac{\chi^2_{n_0-1}}{\sigma_1^2 + \sigma_2^2} \leq \frac{(n_0-1)z}{\sigma_1^2 + \sigma_2^2} \right] $$

$$+ \frac{\sigma_1^2 + \sigma_2^2}{z} P \left[ \frac{\chi^2_{n_0+1}}{\sigma_1^2 + \sigma_2^2} > \frac{(n_0-1)z}{\sigma_1^2 + \sigma_2^2} \right] \right\} \right) \right)$$

$$\leq 2P \left[ \frac{\chi^2_{n_0-1}}{\sigma_1^2 + \sigma_2^2} > \frac{(n_0-1)z}{\sigma_1^2 + \sigma_2^2} \right]$$

$$\leq 2P \left[ \frac{\chi^2_{n_0-1}}{\sigma_1^2 + \sigma_2^2} \right] \quad (2.6.35)$$

**Limiting Distribution of $F''$**

In order to develop the limiting distribution of $F''$ as $n_0 \to \infty$, we need the following lemmas, corollary and definition.
Lemma (2.6.36). Let $C = [c_{ij}]$ be a $k \times k$ positive-definite symmetric matrix with inverse $C^{-1} = [c_{ij}^*]$, say. Then $\sum_{i=1}^{k} \sum_{j=1}^{k} c_{ij}^* c_{ij} = k$. 

Proof. By definition of $C^{-1}$ we have

$$\sum_{j=1}^{k} c_{ij}^* c_{jm} = \begin{cases} 0, & i \neq m \\
1, & i = m 
\end{cases} \quad (i, m = 1, 2, \ldots, k).$$

Since $C$ is symmetric, $C^{-1}$ is also symmetric, so

$$\sum_{j=1}^{k} c_{ij} c_{mj} = \begin{cases} 0, & i \neq m \\
1, & i = m 
\end{cases} \quad (i, m = 1, 2, \ldots, k).$$

hence

$$\sum_{i=1}^{k} \sum_{j=1}^{k} c_{ij} c_{ij}^* = k.$$

Definition (2.6.37). Let $Z(n) = [z_{ij}(n)]$, $i, j = 1, \ldots, k$, $n = 1, 2, \ldots$, be a sequence of random matrices. $Z(n)$ converges stochastically to the matrix $B = [b_{ij}]$ if and only if $z_{ij}(n)$ converges stochastically to $b_{ij}$, $i, j = 1, 2, \ldots, k$. (See Anderson (1959, p. 59).)

Lemma (2.6.38). If $X_1, X_2, \ldots$, is a sequence of $k$-dimensional random variables, if $f : \mathbb{R}^k \to \mathbb{R}^m$ is a measurable mapping which is continuous over a set $B \subseteq \mathbb{R}^k$ for which, $P[X \in B] = 1$, and if $X_n \xrightarrow{p} X$, then $f(X_n) \xrightarrow{p} f(X)$.

Proof. See page 104 of Tucker (1967).
Lemma (2.6.39). Let $Z(n) = [z_{ij}(n)]$ be a sequence of random matrices which are positive-definite with probability one, and which converge stochastically to a positive-definite matrix $B = [b_{ij}]$. Then the sequence of random matrices $Z^{-1}(n)$ converge stochastically to $B^{-1} = [b^{-1}_{ij}]$.

Proof. The inverse operation on matrices may be viewed as a continuous map $f : \mathbb{R}^{k^2} \to \mathbb{R}^{k^2}$. Let $f_{ij}$ be the $ij$th coordinate map for $f$ so $f_{ij} : \mathbb{R}^{k^2} \to \mathbb{R}$, that is $b_{ij} = f_{ij}(b_{11}, b_{12}, ..., b_{kk})$. Then $f_{ij}$ is also continuous for each $i, j = 1, 2, ..., k$. Since $Z(n)$ is positive-definite with probability one we also have $z^{-1}_{ij}(n) = f_{ij}(z_{11}(n), z_{12}(n), ..., z_{kk}(n))$ where $f_{ij}$ is continuous and well-defined except over a null set. Now $z_{ij}(n)$ converges stochastically to $b_{ij}$ and hence, by Lemma (2.6.38), $z^{-1}_{ij}(n)$ converges stochastically to $f_{ij}(b_{11}, b_{12}, ..., b_{kk}) = b_{ij}$, therefore $Z^{-1}(n)$ converges stochastically to $B^{-1}$.

Corollary (2.6.40). If $s = (s_{ij})$ is the sample covariance matrix based on $n > p$ observations from the $p$-dimensional multivariate normal distribution $\mathcal{N}_p(\mu, \Sigma)$, $\Sigma = (\sigma_{ij})$, then $s^{-1} = (s_{ij}^{-1})$ converges stochastically to $\hat{\Sigma}^{-1} = (\sigma_{ij}^{-1})$.

Theorem (2.6.41). The limiting distribution of $F''/k-1$ as $n_0 \to \infty$ is $\chi^2_{k-1}(\Delta)$ where $\Delta = (\mu'(\alpha_{rs})\mu)/z(k-1)$, $\mu = (u_1, u_2, ..., u_{k-1})'$.
Proof. The conditional joint distribution of $\eta_2, \ldots, \eta_k$ is $N_{k-1}(\eta, [z \sum_{i=2}^{k} \sum_{j=2}^{k} \sigma_{ij}^2] (\alpha^{rs}))$, hence by Corollary (2.6.40) the limiting unconditional joint density of $\eta_2, \ldots, \eta_k$ is $N_{k-1}(\eta, [z \sum_{i=2}^{k} \sum_{j=2}^{k} \sigma_{ij}^2] (\alpha^{rs}))$ which, by Lemma (2.6.36), is $N_{k-1}(\eta, z(k-1)(\alpha^{rs}))$. Therefore the limiting joint distribution of $(\eta_2/\sqrt{z(k-1)}, \ldots, \eta_k/\sqrt{z(k-1)})$ is $N_{k-1}(\eta/\sqrt{z(k-1)}, (\alpha^{rs}))$. Now

$$
\frac{F''}{k-1} = \frac{1}{z(k-1)} \sum_{r=2}^{k} \sum_{s=2}^{k} \alpha_{rs}^{\prime} \eta_r \eta_s
$$

which is a continuous function of $\eta_2, \ldots, \eta_k$ which therefore is (by Lemma (2.1.16)) distributed in the limit as $X'(\alpha_{rs})X$ where $X \sim N_{k-1}(\eta/\sqrt{z(k-1)}, (\alpha^{rs}))$ which (by Lemma 2.1.17) is $\chi^2_{k-1}(\Delta)$ with noncentrality parameter $\Delta = (\eta'(\alpha_{rs})\eta)/z(k-1)$ since $(\alpha_{rs})(\alpha^{rs}) = I$, which is idempotent.
2.7. Comparison of Procedures $\mathcal{P}_B^1$, $\mathcal{P}_B^2$ and $\mathcal{P}_B^3$

We have developed three procedures for testing (2.4.2) which can guarantee operating characteristic requirements independent of the unknown variances. We now compare these procedures and measure their relative efficiency.

Recall that procedures $\mathcal{P}_B^1$, $\mathcal{P}_B^2$ and $\mathcal{P}_B^3$ use sampling rules $\mathcal{P}(n_0, 1; s_1^2, z_1)$, $\mathcal{P}(n_0, 2k; s_1^2, z_2)$ and $\mathcal{P}(n_0, (k-1)^2; s_{ij}^2, z_{ij})$.

For $\mathcal{P}_B^1$, we have shown in (2.4.13) that

$$0 \leq E(N_{T_1}) - \sum_{i=1}^{k} M_i < \sum_{i=1}^{k} P \left[ \chi^2_{n_0-1} > \frac{(n_0-1)z_1}{\sigma_i^2} \right],$$

where

$$M_i = (n_0+1) \left[ \frac{\chi^2_{n_0-1}}{\sigma_i^2} > \frac{(n_0-1)z_1}{\sigma_i^2} \right] + \frac{\sigma_i^2}{z_1} \left[ \chi^2_{n_0+1} > \frac{(n_0-1)z_1}{\sigma_i^2} \right].$$

For $\mathcal{P}_B^2$, from (2.5.29)

$$0 \leq E(N_{T_2}) - \sum_{i=1}^{k} M_i' < \sum_{i=1}^{k} P \left[ \chi^2_{n_0-1} > \frac{z_2(n_0+2k)(n_0-1)}{\sigma_i^2} \right],$$

where

$$M_i' = (n_0+2k) \left[ \frac{\chi^2_{n_0-1}}{\sigma_i^2} < \frac{z_2(n_0+2k)(n_0-1)}{\sigma_i^2} \right] + \frac{\sigma_i^2}{z_2} \left[ \chi^2_{n_0+1} > \frac{z_2(n_0+2k)(n_0-1)}{\sigma_i^2} \right];$$

while for $\mathcal{P}_B^3$, from (2.6.35), for $k = 2$ we have
\[ 0 \leq E(N_{T_1}) - 2M' \leq 2P \left[ \frac{\chi^2_{n_0 - 1}}{\sigma_1^2 + \sigma_2^2} > \frac{(n_0^2 - 1)z_2}{2^2 + \sigma_2^2} \right] \]

where

\[ M' = (n_0 + 1) P \left[ \frac{\chi^2_{n_0 - 1}}{\sigma_1^2 + \sigma_2^2} < \frac{(n_0^2 - 1)z_2}{2^2 + \sigma_2^2} \right] + \frac{\sigma_1^2 + \sigma_2^2}{z_2^2} \cdot P \left[ \frac{\chi^2_{n_0 + 1}}{\sigma_1^2 + \sigma_2^2} > \frac{(n_0^2 - 1)z_2}{2^2 + \sigma_2^2} \right]. \]

Hence we also have

\[ 0 \leq E(N_{T_1}) - \sum_{i=1}^{k} M_i \leq k \]

\[ (2.7.1) \]

\[ 0 \leq E(N_{T_2}) - \sum_{i=1}^{k} M'_i \leq k \]

which allow us to write \( E(N_{T_1}) = \sum_{i=1}^{k} M_i + \epsilon(\sigma_1, \ldots, \sigma_k, z_1) \) and \( E(N_{T_2}) = \sum_{i=1}^{k} M'_i + \epsilon'(\sigma_1, \ldots, \sigma_k, z_2) \), where \( 0 \leq \epsilon \leq k \) and \( 0 \leq \epsilon' \leq k \).

These facts are useful in considering the limiting behavior of the ratios of total expected sample size when \( z_1 \) and \( n_0 \) are fixed and the \( \sigma_i^2 \) approach various limits.

Let \( (n_1, n_2, \ldots, n_k) \) be some permutation of the integers \( \{1, 2, \ldots, k\} \) and let \( U = \{a_{n_1}, \ldots, a_{n_k}\} \) and \( V = \{a_{n_{j+1}}, \ldots, a_{n_k}\} \). In the following lemmas limits are taken as the elements of \( U \) independently and arbitrarily converge to zero and the elements of \( V \) independently and arbitrarily converge to infinity.

Lemma (2.7.2).

\[
\lim_{N_T} \frac{E(N_{T_1})}{E(N_{T_2})} = \begin{cases} 
\frac{(n_0+1)/(n_0+2k)}{V = \emptyset} \\
\frac{z_2/z_1}{V \neq \emptyset}
\end{cases}
\]
Proof. If \( V \neq \emptyset \), then

\[
\frac{E(N_{T_1})}{E(N_{T_2})} = \lim_{n \to \infty} \frac{\sum_{i=1}^{k} M_i + \epsilon(\sigma_1', \ldots, \sigma_k', z_1)}{\sum_{i=1}^{k} M'_i + \epsilon'(\sigma_1, \ldots, \sigma_k, z_2)}
\]

\[
= \lim_{n \to \infty} \frac{\sum_{i=1}^{k} M_i}{\sum_{i=1}^{k} M'_i}
\]

since \( \lim_{n \to \infty} M'_i = \infty \) and both \( \epsilon(\sigma_1', \ldots, \sigma_k', z_1) \) and \( \epsilon'(\sigma_1, \ldots, \sigma_k, z_2) \) are bounded. Now, letting \( m = n_0 - 1 \),

\[
\lim_{n \to \infty} \frac{E(N_{T_1})}{E(N_{T_2})} = \lim_{n \to \infty} \frac{(m+2) \sum_{i=1}^{k} \mathbb{P}\left[ \frac{X^2}{\sigma_i^2} > \frac{(n_0^2-1)z_1}{\sigma_i^2} \right] + \frac{1}{z_1} \sum_{i=1}^{k} \sigma_i^2 \mathbb{P}\left[ \frac{X^2}{\sigma_i^2} > \frac{(n_0^2-1)z_1}{\sigma_i^2} \right]}{(n_0+2k) \sum_{i=1}^{k} \mathbb{P}\left[ \frac{X^2}{\sigma_i^2} < \frac{(n_0^2+2k)z_2}{\sigma_i^2} \right] + \frac{1}{z_2} \sum_{i=1}^{k} \sigma_i^2 \mathbb{P}\left[ \frac{X^2}{\sigma_i^2} < \frac{(n_0^2+2k)z_2}{\sigma_i^2} \right]}
\]

\[
= \lim_{n \to \infty} \frac{(m+2) \sum_{i=1}^{k} \mathbb{P}\left[ \frac{X^2}{\sigma_i^2} < \frac{(n_0^2-1)z_1}{\sigma_i^2} \right] + \frac{1}{z_1} \sum_{i=1}^{k} \sigma_i^2 \mathbb{P}\left[ \frac{X^2}{\sigma_i^2} < \frac{(n_0^2-1)z_1}{\sigma_i^2} \right]}{(n_0+2k) \sum_{i=1}^{k} \mathbb{P}\left[ \frac{X^2}{\sigma_i^2} < \frac{(n_0^2+2k)z_2}{\sigma_i^2} \right] + \frac{1}{z_2} \sum_{i=1}^{k} \sigma_i^2 \mathbb{P}\left[ \frac{X^2}{\sigma_i^2} < \frac{(n_0^2+2k)z_2}{\sigma_i^2} \right]}
\]

Since \( \mathbb{P}\left[ \frac{X^2}{n_0+1} < \frac{(n_0^2-1)z_1}{\sigma_i^2} \right] = \int_{0}^{\frac{(n_0^2-1)z_1}{\sigma_i^2}} f(x) \, dx \), where \( f(x) \) is the density of a \( \chi^2_{n_0-1} \) variate, by the mean value theorem for integrals
\[ P \left[ \frac{\chi^2_{n_0+1} \leq (n_0^2-1)z_1}{\sigma_1^2} \right] = \frac{(n_0^2-1)z_1}{\sigma_1^2} f(\xi_1) \] where \( 0 \leq \xi_1 \leq \frac{(n_0^2-1)z_1}{\sigma_1^2} \).

Similarly,

\[ P \left[ \frac{\chi^2_{n_0+1} \leq \frac{(n_0+2k)(n_0-1)z_2}{\sigma_1^2}}{\sigma_1^2} \right] = \frac{(n_0+2k)(n_0-1)z_2}{\sigma_1^2} f(\eta_1), \]

where

\[ 0 \leq \eta_1 \leq \frac{(n_0 + 2k)(n_0-1)z_2}{\sigma_1^2}. \]

Therefore,

\[
\lim_{r_1} \frac{E(N_{r_1})}{E(N_{r_2})} = \lim_{r_1} \frac{\sum_{i=1}^{k} P \left[ \frac{\chi^2_{m} \leq \frac{(n_0^2-1)z_1}{\sigma_1^2}}{\sigma_1^2} \right] + \frac{1}{z_1} \left[ \sum_{i=1}^{k} \sigma_i^2 - \sum_{i=1}^{k} \frac{(n_0^2-1)z_1}{\sigma_1^2} f(\xi_1) \right]}{\left( n_0 + 2k \right) \sum_{i=1}^{k} P \left[ \frac{\chi^2_{m} \leq \frac{(n_0+2k)mz_2}{\sigma_1^2}}{\sigma_1^2} \right] + \frac{1}{z_2} \left[ \sum_{i=1}^{k} \sigma_i^2 - \sum_{i=1}^{k} \frac{(n_0+2k)mz_2}{\sigma_1^2} f(\xi_1) \right]}.
\]

Since

\[ V \neq \emptyset, \sum_{i=1}^{k} \sigma_i^2 \to \infty, (n_0+1) \sum_{i=1}^{k} P \left[ \frac{\chi^2_{n_0-1} \leq \frac{(n_0^2-1)z_1}{\sigma_1^2}}{\sigma_1^2} \right]. \]

and
are bounded, and hence dividing each term by \( \sum_{i=1}^{k} \sigma_i^2 \) we obtain

\[
\lim E\left( \frac{N_1}{N_2} \right) = \left( \frac{1/z_1}{1/z_2} \right) = \frac{z_2}{z_1}.
\]

If \( V = \emptyset \) then all the variances are converging to zero and only the initial sample size (perhaps +1) is needed per population and thus

\[
\lim E\left( \frac{N_1}{N_2} \right) = \left( \frac{n_0+1}{n_0+2k} \right).
\]

For \( k = 2 \) we can compare \( \mathcal{P}_{B_1} \) with \( \mathcal{P}_{B_3} \), and \( \mathcal{P}_{B_2} \) with \( \mathcal{P}_{B_3} \). Letting \( U \) and \( V \) be defined as above we have the following lemmas.

**Lemma (2.7.3).**

\[
\lim E\left( \frac{N_1}{N_2} \right) = \begin{cases} 
\frac{z_2}{z_1}, & V \neq \emptyset \\
1, & V = \emptyset .
\end{cases}
\]

**Proof.** The proof is similar to that of Lemma 2.7.2.
Lemma (2.7.4).

\[ \lim_{t \to 2} \frac{E(N_t)}{E(N_{T_2})} = \begin{cases} 
\frac{z_3}{2z_2}, & \forall \neq \emptyset \\
\frac{(n_0+2)}{(n_0+1)}, & \forall = \emptyset 
\end{cases} \]

Proof. As in Lemma (2.7.2).

Table 2.7.5 presents the limiting values of

a) \( E(N_{T_1})/E(N_{T_2}) \)

b) \( E(N_{T_1})/E(N_{T_3}) \)

c) \( E(N_{T_2})/E(N_{T_3}) \)

for \( k = 2 \) and the various limiting conditions.

<table>
<thead>
<tr>
<th>( \sigma_1 \to 0, \sigma_2 \to 0 )</th>
<th>( \sigma_1 \to 0, \sigma_2 \to \infty ) or ( \sigma_1 \to \infty, \sigma_2 \to 0 ) or ( \sigma_1 \to \infty, \sigma_2 \to \infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) ( (n_0+1)/(n_0+2) )</td>
<td>a) ( \frac{z_2}{z_1} )</td>
</tr>
<tr>
<td>b) ( (n_0+1)/(n_0+1) = 1 )</td>
<td>b) ( \frac{z_3}{2z_1} )</td>
</tr>
<tr>
<td>c) ( (n_0+2)/n_0 + 1 )</td>
<td>c) ( \frac{z_3}{2z_2} )</td>
</tr>
</tbody>
</table>
The above comparisons shed little light on the relative efficiencies of the three procedures in that the $z_1$ should be set to make the tests have equal power. At the present time the exact distributions of $\mathcal{F}'$ and $\mathcal{F}''$ have not been simplified enough to evaluate $z_2$ and $z_3$ analytically or numerically for a fixed power. Also the case $k = 2$ probably is not typical for $\mathcal{P}_{B_3}$ since as $k$ increases $n_0^+ (k-1)^2$ increases rapidly. Thus $\mathcal{P}_{B_3}$ requires a much greater number of initial observations than $\mathcal{P}_{B_1}$ or $\mathcal{P}_{B_2}$ as $k$ increases. Some insight into the relative behavior of the procedures can be gained if we look at the limiting distributions since all three are chi-square. The remaining part of this section deals with the asymptotic comparison of $\mathcal{P}_{B_1}$, $\mathcal{P}_{B_2}$ and $\mathcal{P}_{B_3}$.

Let $\chi^2 \sim \chi^2_{k-1}(\Delta)$ and let $\Delta_\beta$ denote the noncentrality needed to achieve $\beta = P[\chi^2 > \chi^2_{k-1; \Delta}].$ If we use $\tilde{F}$ to test (2.4.2) at level $\alpha$ we want $\beta = P[\tilde{F} > \chi^2_{k-1; \Delta_1}],$ where $\Delta_1 = \sum_{i=1}^k (\mu_i - \bar{\mu})^2 / z_1.$ Thus if $\Delta = \sum_{i=1}^k (\mu_i - \bar{\mu})^2$ we then choose $z_1$ so that $\Delta / z_1^2 = \Delta_\beta$ or $z_1 = \Delta / \Delta_\beta.$

If we use $F'$ to test (2.4.2) we then want

$$\beta = P[F'/k > \chi^2_{k-1; \Delta_2}]$$

where $\Delta_2 = \Delta / k z_2,$ so $\Delta / k z_2 = \Delta_\beta$ or $z_2 = \Delta / k \Delta_\beta.$

In general the limiting distribution of $\mathcal{F}''$ is a noncentral chi-square with $k-1$ degrees of freedom but its noncentrality parameter is not of the same form as that of $\tilde{F}$ or $F'$ except when $k = 2.$ In that case we want $\beta = P[F'' > \chi^2_{1; \Delta_3}]$ where $\Delta_3 = (\mu_1 - \mu_2)^2 / z_3,$ so $z_3 = (\mu_1 - \mu_2)^2 / \Delta_\beta.$
We are now in a position to make the following asymptotic comparisons. The limiting value of \( z_2/z_1 \) in Lemma (2.7.2) is
\[
z_2/z_1 = (\Delta/k\Delta_p)/(\Delta/\Delta_p) = 1/k.
\]
For \( k = 2 \), \( \Delta = \sum_{i=1}^{k} (\mu_i - \mu)^2 = (\mu_1 - \mu_2)^2/2 \)
and the limiting value of \( z_2/2z_1 \) in Lemma (2.7.3) becomes \( z_2/2z_1 = [(\mu_1 - \mu_2)^2/\Delta_p]/[2(\mu_1 - \mu_2)^2/2\Delta_p] = 1 \). The corresponding value of
\[
z_3/2z_2 \text{ for the comparison of } \mathcal{P}_{B_2} \text{ with } \mathcal{P}_{B_3} \text{ becomes}
\]
\[
z_3/2z_2 = [(\mu_1 - \mu_2)^2/\Delta_p]/[2(\mu_1 - \mu_2)^2/2\Delta_p] = 2.
\]
Thus for \( k = 2 \) we can construct a table similar to (2.7.5) which presents the limiting values of the ratios of expected sample sizes. The values are found in Table 2.7.6.

### Table 2.7.6. Limiting Values of \( a) E(N_{T_1})/E(N_{T_2}), \) \( b) E(N_{T_1})/E(N_{T_3}), \) and \( c) E(N_{T_2})/E(N_{T_3}) \) for \( k = 2. \)

<table>
<thead>
<tr>
<th>( V )</th>
<th>( V \neq 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) ( n_0+1)/(n_0+2) )</td>
<td>a) ( z_2/z_1 = 1/2 )</td>
</tr>
<tr>
<td>b) ( n_0+1)/(n_0+1) = 1 )</td>
<td>b) ( z_3/2z_1 = 1 )</td>
</tr>
<tr>
<td>c) ( n_0+2)/(n_0+1) )</td>
<td>c) ( z_2/2z_2 = 2 )</td>
</tr>
</tbody>
</table>

From Table 2.7.6 we see that if all \( \sigma_i^2 \to 0 \) the three procedures are essentially equivalent for \( k = 2 \) differing by at most one observation per population. If any \( \sigma_i^2 \to \infty \) then \( \mathcal{P}_{B_2} \) requires twice as many observations as \( \mathcal{P}_{B_1} \) and \( \mathcal{P}_{B_2} \) while \( \mathcal{P}_{B_1} \) and \( \mathcal{P}_{B_3} \) require the same number of observations.
We now consider the behavior of these ratios as the power $\beta$ of the test converges to one with alternative $\Delta$ fixed. Again we compare $P_{B_1}$ and $P_{B_2}$ for general $k$, and $P_{B_1}$ with $P_{B_3}$ and $P_{B_2}$ with $P_{B_3}$ for $k = 2$, in terms of limiting distributions.

Lemma (2.7.7). $\lim_{\beta \to 1} \frac{E(N_{T_1})}{E(N_{T_2})} = l/k$ for $\Delta$ fixed.

Proof. Let $m = n_0 - 1$ then since $\Delta \to \infty$, $z_1 \to 0$ and $z_2 \to 0$ as $\beta \to 1$. Thus

$$\lim_{\beta \to 1} \frac{E(N_{T_1})}{E(N_{T_2})} = \frac{\sum_{i=1}^{k} M_i}{\sum_{i=1}^{k} M'_i}$$

$$= \lim_{\beta \to 1} \frac{\sum_{i=1}^{k} M_i}{\sum_{i=1}^{k} M'_i}$$

$$= \lim_{\beta \to 1} \frac{(m+2) \sum_{i=1}^{k} P \left[ \chi_{m}^{2} < \frac{(n_0^2-1)z_1}{\sigma_i^2} \right] + \frac{1}{z_1} \sum_{i=1}^{k} \sigma_i^2 P \left[ \chi_{m+2}^{2} > \frac{(n_0^2-1)z_1}{\sigma_i^2} \right]}{(n_0+2k) \sum_{i=1}^{k} P \left[ \chi_{m}^{2} < \frac{(n_0+2k)z_2}{\sigma_i^2} \right] + \frac{1}{z_2} \sum_{i=1}^{k} \sigma_i^2 P \left[ \chi_{m+2}^{2} > \frac{(n_0+2k)z_2}{\sigma_i^2} \right]}$$
However \( P[X_m^2 < (n_0^2 - 1)z_1^2/\sigma_i^2] \) and \( P[X_m^2 < (n_0 + 2k)m_2^2/\sigma_i^2] \to 0, \)

while \( P[X_m^2 > (n_0^2 - 1)z_1^2/\sigma_i^2] \) and \( P[X_m^2 > (n_0 + 2k)m_2^2/\sigma_i^2] \to 1, \)

so

\[
\lim_{\beta \to 1} \frac{E(N_{T_1})}{E(N_{T_2})} = \lim_{\beta \to 1} \frac{\frac{1/z_1^2}{1/z_2^2}}{\frac{1/z_1^2}{1/z_2^2}} = \lim_{\beta \to 1} \frac{z_2/z_1}{z_2/z_1} = \lim_{\beta \to 1} \frac{\Delta/\kappa_{\beta}}{\Delta/\kappa_{\beta}} = 1/k.
\]

Lemma (2.7.8). For \( k = 2, \lim_{\beta \to 1} E(N_{T_1})/E(N_{T_3}) = 1. \)

Proof. The proof follows as in Lemma (2.7.7).

Lemma (2.7.9). For \( k = 2, \lim_{\beta \to 1} E(N_{T_1})/E(N_{T_3}) = 2. \)

Proof. The proof follows as in Lemma (2.7.7). Note also that

\[
\lim_{\beta \to 1} E(N_{T_1})/E(N_{T_j}) \text{ for a fixed alternative } \Delta \text{ is the same as }
\]

\[
\lim_{\Delta \to 0} E(N_{T_1})/E(N_{T_j}) \text{ for a fixed power } \beta.
\]

While the above comparisons are not comprehensive they do indicate that: \( P_{B_1} \) is superior to both \( P_{B_2} \) and \( P_{B_3} \) (for the same power at a fixed alternative it will require fewer observations on the average). Also \( P_{B_1} \) is very simple to implement and to program for a computer. As mentioned before \( P_{B_3} \) might be advantageous if we are looking at a situation where one group may be considered as a
control. However, \( P_{B_3} \) does require randomization, which has been criticized by some authors. Overall we presently would recommend use of \( P_{B_1} \).
2.8. Comparison of Procedures Based on $\tilde{X}$'s with Procedures Based On $\bar{X}$'s.

A natural question is what happens to our testing and multiple-comparison procedures if we replace the generalized sample means $\tilde{X}_i$ with the sample means $\bar{X}_i$. Intuitively one might expect the $\tilde{X}_i$ modification to perform better. However, Rinott has answered a problem of Dudewicz and Dalal and shown that, in the case of selection problems, it is not true that the $\tilde{X}_i$ modification performs uniformly better than the original $\bar{X}$-based procedure.

We consider this question for the case $k = 2$, namely the Behrens-Fisher context, and study the power functions of the tests based on $\tilde{X}_i$ and $\bar{X}_i$, as well as the corresponding confidence intervals generated for the difference $\mu_1 - \mu_2$. Our analysis utilizes results of Ruben presented in Section 1.3. For ease of reference we denote by

$$\tilde{P}_B \left[ P(n_0; l, s_i^2, z) \right]$$ to $\tilde{X}_i$ and by $\bar{P}_B \left[ P(n_0; l, s_i^2, z) \right]$ to $\bar{X}_i$ but uses $\bar{X}_i$ the overall sample mean. Let $X = (\bar{X}_1, \bar{X}_2)$ be the vector of sample means. The (exact) probability distribution of $\bar{X}$ for any subset $R$ of $\mathbb{R}^2$ is

$$P[\bar{X} \in R|\mu, \sigma] = \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \frac{p_{m_1}(\sigma)}{m_1!} \frac{p_{m_2}(\sigma_2)}{m_2!} Q_m(R|\mu, \sigma)$$

where $\mu = (\mu_1, \mu_2)$, $\sigma = (\sigma_1, \sigma_2)$, $p_{m_i}(\sigma_i) = P[N_i = m_i|\sigma_i]$
\[ Q_m(R|\mu, \Sigma_m) = P[\mathbf{X} \in R|n=m; \mu, \sigma] = (2\pi)^{-1} \int \int_{\mu+U_{m}^{1/2} \xi \in R} e^{-\frac{1}{2} \xi' \Sigma_m^{-1} \xi} \, d\xi \]

\[ n = (N_1, N_2), \quad m = (m_1, m_2), \quad \Sigma_m = \begin{bmatrix} \sigma_1^2/m_1 & 0 \\ 0 & \sigma_2^2/m_2 \end{bmatrix} \]

and \( \xi \sim N_2(0, I) \). Ruben (see Ruben (1962a), p. 160-162) proved that for any subset \( R \) of \( \mathbb{R}^2 \),

\[ \lim_{\sigma \to \infty} P[\mathbf{X} \in R|\mu, \sigma] = \int_{0}^{\infty} \int_{0}^{\infty} Q^*(R; \mu, U) \, dG \quad (2.8.1) \]

where \( G \) is the joint distribution function of two independent \( \chi^2_{n_0-1} \) variates \( u_1 \) and \( u_2 \), and

\[ Q^* = (2\pi)^{-1} \int \int_{\mu+U_{\xi} \xi \in R} e^{-\frac{1}{2} \xi' \Sigma^{-1} \xi} \, d\xi \]

where

\[ U = \begin{bmatrix} (n_0-1) \sqrt{v}/\sqrt{u_1} & 0 \\ 0 & (n_0-1) \sqrt{v}/\sqrt{u_2} \end{bmatrix} \]

The limit is taken as \( \sigma_1 \) and \( \sigma_2 \) independently and arbitrarily converge to infinity. Thus in the limit \( \mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2) \) is distributed as \( (\mu_1 + \sqrt{v} t_1, \mu_2 + \sqrt{v} t_2) \) where \( t_1 \) and \( t_2 \) are i.i.d. \( t_{n_0-1} \) variates, which is exactly how \((\mathbf{X}_1, \mathbf{X}_2)\) is distributed for all values of \((\sigma_1, \sigma_2)\).
Ruben also showed that

\[
\lim_{\sigma \to \infty} P[\bar{X} \in R|\mu, \sigma] = \left\{
\begin{array}{ll}
\inf_{\sigma} P[\bar{X} \in R|\mu, \sigma], & Q_m \text{ strictly decreasing in } \sigma_1^2 \text{ and } \sigma_2^2; \\
\sup_{\sigma} P[\bar{X} \in R|\mu, \sigma], & Q_m \text{ strictly increasing in } \sigma_1^2 \text{ and } \sigma_2^2.
\end{array}
\right.
\] (2.8.2)

When \( k = 2 \) we have already shown (see equation (2.4.8)) that

\[
\tilde{F} = \sum_{i=1}^{2} (\tilde{x}_1 - \tilde{x})^2/z = (\tilde{x}_1 - \tilde{x}_2)^2/(2z).
\]

To test \( H_0: \mu_1 = \mu_2 \) we reject if and only if

\[
\frac{|\tilde{x}_1 - \tilde{x}_2|}{\sqrt{z}} > \sqrt{2} \, \alpha/2,
\]

where \( \alpha/2 \) is the upper \( \alpha \)th percent point of the distribution of the difference of two independent \( t_{n_0-1} \) variates. Letting \( \eta = \mu_1 - \mu_2 \), the power of this test is

\[
\hat{\beta}(\eta) = 1 - P \left[ -\sqrt{2z} \, \alpha/2 - \eta \leq \frac{\tilde{x}_1 - \tilde{x}_2}{\sqrt{z}} < \sqrt{2z} \, \alpha/2 - \eta \right]
\]

\[
= 1 - P \left[ -\sqrt{2z} \, \alpha/2 - \eta \leq t_{1-t_2} \leq \sqrt{2z} \, \alpha/2 - \eta \right]. \quad (2.8.3)
\]
If we replace $\tilde{X}_1$ with $\tilde{X}_1$ then we are basing our test on the statistic $F = \frac{\sum_{i=1}^{2} (\tilde{X}_i - \bar{X})^2}{z}$ where $\bar{X} = (\tilde{X}_1 + \tilde{X}_2)/2$ and so $F = (\tilde{X}_1 - \tilde{X}_2)^2/(2z)$ and the power of this test is given by

$$\beta(\eta) = 1 - P[-\sqrt{2z} \, d\alpha/2 - \eta \leq \tilde{X}_1 - \tilde{X}_2 \leq \sqrt{2z} \, d\alpha/2 - \eta]. \quad (2.8.4)$$

From equation (2.8.1) we see that the limiting power of the test based on $\tilde{X}_1$ is the same as the power of the test based on $\tilde{X}_1$. The non-limiting power, however, depends upon the unknown variances. Whether it is larger or smaller than the limiting power (and hence the power of the test based on $\tilde{X}_1$) depends upon whether $Q_m(R/\mu, \sigma)$ is increasing or decreasing in $\sigma_1$ and $\sigma_2$ with

$$R = \left\{ (\tilde{X}_1, \tilde{X}_2) : \left| \frac{\tilde{X}_1 - \tilde{X}_2}{\sqrt{z}} \right| > \sqrt{2} \, d\alpha/2 \right\}.$$ 

If $Q_m$ is decreasing in $\sigma_1^2$ and $\sigma_2^2$ then the non-limiting power is larger while if it is increasing in $\sigma_1^2, \sigma_2^2$ the non-limiting power is smaller than the power given by the limiting distribution. Now

$$Q_m(R/\mu, \sigma) = 1 - P[-\sqrt{2} \, d\alpha/2 \leq (\tilde{X}_1 - \tilde{X}_2)/\sqrt{z} \leq \sqrt{2} \, d\alpha/2]$$

$$= 1 - P \left[ -\sqrt{\frac{2z}{\sigma_1^2/m_1 + \sigma_2^2/m_2}} \leq Y \leq \frac{\sqrt{2z} - \eta}{\sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}} \right] \quad (2.8.5)$$

where $Y \sim N(0,1)$,
since \( Q_m \) is the conditional probability distribution for \((\bar{X}_1, \bar{X}_2)\) given \( N_1=m_1 \) and \( N_2=m_2 \), which is \( N_2(\mu_m, 1) \). We wish to consider (2.8.5) individually in \( \sigma_1^2 \) and \( \sigma_2^2 \), and by symmetry we need only consider one of them, say \( \sigma_1^2 \). Considering (2.8.5) as a function of \( \sigma_1^2 \), we may write

\[
\xi(\sigma_1^2) = 1 - \Phi \left( \frac{\sqrt{2z} \, d\alpha/2 - \eta}{\sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}} \right) + \Phi \left( \frac{-\sqrt{2z} \, d\alpha/2 - \eta}{\sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}} \right)
\]

Then

\[
\frac{d\xi(\sigma_1^2)}{d\sigma_1^2} = -\Phi \left[ \frac{\sqrt{2z} \, d\alpha/2 - \eta}{\sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}} \right] \left[ -\frac{1}{2} \left( \frac{\sqrt{2z} \, d\alpha/2 - \eta}{m_1} \right) \left( \frac{\sigma_1^2 + \sigma_2^2}{m_1} \right)^{-3/2} \right]
\]

\[
+ \Phi \left[ \frac{-\sqrt{2z} \, d\alpha/2 - \eta}{\sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}} \right] \left[ -\frac{1}{2} \left( \frac{-\sqrt{2z} \, d\alpha/2 - \eta}{m_1} \right) \left( \frac{\sigma_1^2 + \sigma_2^2}{m_2} \right)^{-3/2} \right],
\]

which is \(< 0\) if and only if

\[
\Phi \left[ \frac{\sqrt{2z} \, d\alpha/2 - \eta}{\sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}} \right] (\sqrt{2z} \, d\alpha/2 - \eta)
\]

\[
< \Phi \left[ \frac{-\sqrt{2z} \, d\alpha/2 - \eta}{\sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}} \right] (-\sqrt{2z} \, d\alpha/2 - \eta)
\]

(2.8.6)
If $|\eta| < \sqrt{2} \alpha/2$, then $\sqrt{2} \alpha/2 - \eta > 0$ and $-\sqrt{2} \alpha/2 - \eta < 0$, so (2.8.6) never holds, hence $Q_m$ is increasing in $\sigma_1^2$ and $\sigma_2^2$.

If $\eta > \sqrt{2} \alpha/2$, then $d\sigma_1^2/d\sigma_1 < 0$ if and only if

$$
\phi \left[ \frac{\sqrt{2} \alpha/2 - \eta}{\sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}} \right] > \phi \left[ \frac{-\sqrt{2} \alpha/2 - \eta}{\sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}} \right]
$$

if and only if

$$
\exp \left\{ -\frac{1}{2} \left[ \frac{(\sqrt{2} \alpha/2 - \eta)^2}{\sigma_1^2/m_1 + \sigma_2^2/m_2} \right] + \frac{1}{2} \left[ \frac{(-\sqrt{2} \alpha/2 - \eta)^2}{\sigma_1^2/m_1 + \sigma_2^2/m_2} \right] \right\} > \frac{(-\sqrt{2} \alpha/2 - \eta)}{(\sqrt{2} \alpha/2 - \eta)}
$$

if and only if

$$
\frac{1}{2(\sigma_1^2/m_1 + \sigma_2^2/m_2)} \left[ (-\sqrt{2} \alpha/2 - \eta)^2 - (\sqrt{2} \alpha/2 - \eta)^2 \right]
$$

$$
> \log \frac{(-\sqrt{2} \alpha/2 - \eta)}{(\sqrt{2} \alpha/2 - \eta)}
$$

if and only if

$$
\frac{\sigma_1^2}{m_1} + \frac{\sigma_2^2}{m_2} < \frac{2(\sqrt{2} \alpha/2)\eta}{\log \left( \frac{-\sqrt{2} \alpha/2 - \eta}{\sqrt{2} \alpha/2 - \eta} \right)}, \quad (2.8.7)
$$
hence we see that if $\eta > \sqrt{2z} \frac{d\alpha}{2}$ then $g(\sigma_1^2)$ is strictly decreasing in $\sigma_1^2$ as long as (2.8.7) holds. Similarly it can be shown that if $\eta < -\sqrt{2z} \frac{d\alpha}{2}$ then $g(\sigma_1^2)$ is strictly decreasing in $\sigma_1^2$ as long as (2.8.7) holds.

Thus we see that if $|\eta| < \sqrt{2z} \frac{d\alpha}{2}$ the true power curve lies below the limiting power curve, while if $|\eta| > \sqrt{2z} \frac{d\alpha}{2}$ and the variances are sufficiently small then the true power curve lies above the limiting curve hence the procedure based on $\bar{X}_1$ is not uniformly better than the procedure based on $\bar{X}_1$. A typical graph of the two power functions would be as in Figure 2.8.8 with the point of intersection depending upon the variances.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.8.8.png}
\caption{FIGURE 2.8.8.}
\end{figure}
Note that the test based on \( \tilde{X}_1 \) and the test based on \( \tilde{X}_1 \) with rejection regions \( \left| \frac{(\tilde{X}_1 - \tilde{X}_2)}{\sqrt{2}} \right| > \sqrt{2} \alpha_2/2 \) and \( \left| \frac{(\tilde{X}_1 - \tilde{X}_2)}{\sqrt{2}} \right| > \sqrt{2} \alpha_2/2 \) respectively, have the same level (since the level of the test is the sup over all nuisance parameters of the probability of acceptance, see page 61 of Lehmann (1959)) and hence are comparable.

This analysis also shows that Ruben's procedure only controls the power at a given alternative asymptotically since the true power is greater or smaller than the limiting power depending upon the size of the variances.

Suppose we now consider a confidence interval for \( \eta = \mu_1 - \mu_2 \) based on \( \tilde{X}_1 \) or \( \tilde{X}_1 \). A confidence interval for \( \eta \) based on \( \tilde{X}_1 \) would depend upon the probability

\[
P \left[ -\frac{\tilde{X}_1 - \tilde{X}_2 - \eta}{\sqrt{2}} \leq \frac{\alpha}{2} \leq \frac{\alpha}{2} \right] = 1 - \alpha
\]

hence the interval \( \{ \tilde{X}_1 - \tilde{X}_2 + \sqrt{2} \alpha/2 \} \) would have confidence coefficient \( 1 - \alpha \). The corresponding interval based on \( \tilde{X}_1 \) would be

\( \{ \tilde{X}_1 - \tilde{X}_2 + \sqrt{2} \alpha/2 \} \) which has a confidence coefficient of \( 1 - \alpha \) as its limiting value. The true confidence coefficient is greater (less) than \( 1 - \alpha \) depending on whether \( Q_m = P[ -\frac{\alpha}{2} \leq (\tilde{X}_1 - \tilde{X}_2 - \eta)/\sqrt{2} < \frac{\alpha}{2} ] \) is decreasing (increasing) in \( \sigma_1^2, \sigma_2^2 \). Now

\[
Q_m = P \left[ -\frac{\sqrt{2} \alpha/2}{\sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}} \leq Y \leq \frac{\sqrt{2} \alpha/2}{\sqrt{\sigma_1^2/m_1 + \sigma_2^2/m_2}} \right] , \quad Y \sim N(0,1)
\]
which is clearly decreasing in both $\sigma_1^2$ and $\sigma_2^2$. Thus the true confidence coefficient is at least $1 - \alpha$ and so we do better using $\bar{X}_i$ rather than $\widetilde{X}_i$. (Note it has been shown in other problems that there is not a uniform gain using $\bar{X}_i$'s and even when there is it is of a trivial nature. It is not obvious here what the magnitude of the gain is and it merits further study.)
2.9. A Note on the Transformation of One-Stage Procedures into Two-Stage Procedures.

Statisticians are sometimes confronted by clients who already have their data in hand, and who do not wish to take a second sample. In those situations two-stage procedures such as considered above are not directly feasible. For such situations it has often been suggested to us, after talks at many institutions, that it should be possible to transform two-stage procedures into single-stage procedures by appropriate adjustment of the desired operating characteristics.

We now investigate these suggestions in the context of Stein's confidence interval of length \( l \) and confidence coefficient \( 1 - \alpha \) (independent of \( \sigma^2 \)) for the mean \( \mu \) of a single normal population \( \pi \sim N(\mu, \sigma^2) \). This interval is generated by applying \( P(n_0, 1, s^2, z) \) to \( \pi_1 \) with \( s^2 \) being the usual unbiased estimator of \( \sigma^2 \) based on the first \( n_0 \) observations and with \( z = l^2/(4t^2_{\alpha/2}) \). The adjustment method most often suggested, when put into this context is the following.

Method (2.9.1). The \( \alpha \)-Adjustment Method (for Avoiding a Second-Stage).

If a one-stage sample of size \( n \), \((X_1, \ldots, X_n)\) has been taken on \( \pi \), generate a confidence interval for \( \mu \) of length \( l \) in the following manner. Let \( n-1 = n_0 \) in Stein's two-stage procedure and take the first \( n-1 \) observations as the initial sample. Then set \( z \) (i.e. \( \alpha = \alpha(s^2) \) since \( z = l^2/(4t^2_{\alpha/2}) \)) so that

\[
n = \max\{n, \lfloor s^2/z \rfloor + 1\} \tag{2.9.2}
\]
and choose constants \(\{a_i\}\) such that

\[
a_1 = \cdots = a_{n-1}
\]

\[
s^2 \sum_{i=1}^{n} a_i^2 = z \quad (2.9.3)
\]

\[
\sum_{i=1}^{n} a_i = 1.
\]

Let \(\bar{X} = \sum_{i=1}^{n} a_i X_i\), and take as the interval

\[
\bar{X} \pm \ell/2
\]

and claim it has confidence coefficient \(1 - \alpha(s^2)\).

We analyze Method (2.9.1) in Theorem (2.9.4).

**Theorem (2.9.4).** The confidence interval \(I\) generated by Method (2.9.1) actually has confidence coefficient \(1-\alpha\) (say) such that \(1-\alpha = 2\Phi[\ell \sqrt{c/2s}] - 1\), which may range from 0 to 1.

**Proof.** Fix \(c\) such that \(0 < c < n\) and set \(z = s^2/c\) so that (2.9.2) is satisfied. Take \(s^2 \sum_{i=1}^{n} a_i^2 = z\) implies \(\sum a_i^2 = l/c\) and constants \(a_1, \ldots, a_n\) exist satisfying (2.9.3). In fact we may take \(a_1 = \cdots = a_{n-1} = a\) and \(a_n = b\) where

\[
b = \frac{1}{n} \left(1 + \sqrt{(n-1)(n/c-1)}\right)
\]

and

\[
a = \frac{1-b}{n}.
\]
The constants \( \{a_i\} \) are independent of \( s^2 \) and since \( a_1 = \cdots = a_{n-1} \), \( \sum a_i X_i \) is independent of \( s^2 \), therefore unconditionally, \( \tilde{X} \sim N(\mu, \frac{s^2}{c}) \).

If we take \( \tilde{X} \pm \frac{\ell}{2} \) as our confidence interval then, letting \( Y \sim N(0,1) \),

\[
1 - \alpha = P\left[-\frac{\ell}{2} \leq \tilde{X} - \mu \leq +\frac{\ell}{2}\right]
\]

\[
= P\left[-\frac{\ell}{2} \frac{\sqrt{c}}{2\sigma} \leq Y \leq \frac{\ell}{2} \frac{\sqrt{c}}{2\sigma}\right]
\]

\[
= \Phi\left(\frac{\ell}{2} \frac{\sqrt{c}}{2\sigma}\right) - \Phi\left(-\frac{\ell}{2} \frac{\sqrt{c}}{2\sigma}\right)
\]

\[
= 2[\Phi\left(\frac{\ell}{2} \frac{\sqrt{c}}{2\sigma}\right) - 1/2].
\]

From this it is clear that

\[
\lim_{\sigma \to 0} (1 - \alpha) = 0
\]

and

\[
\lim_{\sigma \to \infty} (1 - \alpha) = 1
\]

It should be noted that even conditionally given \( s^2 \) the confidence coefficient for the interval \( I \) is \( 2[\Phi(\ell \sqrt{c/2\sigma}) - 1/2] \) and not \( 1 - \alpha(s^2) \). This follows from the fact that \( \tilde{X} \) and \( s^2 \) are independent.

Corollary (2.9.5). \( \sup_c (1 - \alpha) = 2[\Phi(\ell \sqrt{n/2\sigma}) - 1/2] \) which is the same as the confidence coefficient for the interval \( \bar{X} \pm \frac{\ell}{2} \) where \( \bar{X} \) is the sample mean based on all \( n \) observations.
That the choice of \( n_0 \) (\( 1 \leq n_0 \leq n \)) is of no consequence can be seen as follows. Let \( m < n-1 \), set \( n_0 = m \) and choose \( z \) so that

\[
n = \max\{m+1,[s^2/z]+1\}
\]

This implies \( z = s^2/n \) (or \( c = n \)) and the previous results for this choice of \( c \) holds.

Other methods of adjustment which have been proposed and require a study similar to the above include the following.

**Method (2.9.6). The \( l \)-Adjustment Method (for Avoiding a Second Stage).**

This method is similar to (2.9.1) except that, instead of adjusting the confidence coefficient so that (2.9.2) holds, the experimenter adjusts the length \( l \) of the confidence interval.

**Method (2.9.7). The \( \alpha \)-Level Method (for "Using the Whole First Stage").**

In some experiments it may happen that the choice of \( n_0 \) was "larger than needed to attain a 1-\( \alpha \) confidence coefficient." For example, if \( n_0 = 12 \) and \( s^2/z = 8 \), the two-stage procedure actually throws away information to achieve an exact 1-\( \alpha \) confidence coefficient. One then adjusts 1-\( \alpha \) so that \([s^2/z] = 12\).

That this adjusted 1-\( \alpha \) confidence coefficient is the true confidence coefficient is doubtful and further study on this procedure is needed.
Finally an investigation of methods such as (2.9.1), (2.9.6), and (2.9.7) is also needed for the cases of more than one population and for other goals, for example, testing (power) and ranking and selection (P[CS]). Note that additional complexities arise when \( k \geq 2 \), since \( 1-\alpha \) (e.g.) may need to be forced to be the smallest of \( k \) quantities, resulting in a grossly inefficient procedure.
CHAPTER III

MONTE CARLO SAMPLING STUDIES

3.1. Design of the Monte Carlo Study

In this chapter we present the results of Monte Carlo sampling experiments designed to calculate the distribution of \( \tilde{F} \) and to study the behavior of testing procedures based on \( F \) and \( \tilde{F} \) in a one-way layout under various configurations of standard deviations \( \sigma = (\sigma_1, \ldots, \sigma_k) \) and sample sizes \( n = (n_1, \ldots, n_k) \) for the \( k \) populations. We have tabled the upper 10\%, 5\% and 1\% points of the null distribution of \( \tilde{F} \) for \( k = 2, 3, 4, 5, 6 \) and \( n_0 = 5, 10, 15, 25 \). The power at these points is also tabled for \( \delta^* = \sum_{i=1}^{k} (\mu_i - \mu)^2 \) values of \( 1, 2, 5 \) and \( z = 1, .8, .6, .4, .2, .1, .08, .06, .04, .02, .01 \). We also check the validity of the approximations, to the true distribution of \( \tilde{F} \), which were suggested in Section 2.4.

A study was made of the robustness of the \( F \)-test to inequality of variances in terms of level and power for various configurations of \( \sigma \) and \( n \). Based on these results a comparison of sample-sizes required by the \( F \)-test and \( B_1 \) to achieve the same power for a fixed level and alternative is made.

All of the programs for these Monte Carlo studies were written in FORTRAN and were run at the Instruction and Research Computer Center.
of the Ohio State University on their IBM 370/168 computer. Listings of the programs, along with samples of the output, may be found in Appendix A. The forms of the tables presented in this chapter are modifications of the tables actually produced by the programs. Only two basic programs are presented in Appendix A. Program I was designed to calculate (by Monte Carlo) \( P[F > c] \) for \( c = 1, 1.25, 20, 75 \), for all combinations of \( k, n_0, z \) and \( \bar{\delta}^* \) previously mentioned. Program II was designed to compare the nominal (i.e. assumed) 10\%, 5\% and 1\% levels of the usual F-test and the limiting approximations with their true levels for various configurations of \( g \) and \( n \) for the F-test and for various configurations of \( g \) with \( n_0 = 10 \) for \( F_{B1}^* \). Power studies of the F-test for these nominal levels were based on the same program modified by adding the appropriate constants \( \mu_i \) to the observations from the \( i \)th population to achieve the desired \( \delta^* \) value.

Each of these Monte Carlo experiments was based on 10,000 replications. Therefore, since if we are estimating a probability \( p \) the standard deviation of our estimate is \( \sqrt{p(1-p)/10000} \), the maximum standard deviation for any of our estimates is \( 1/2 \sqrt{10000} = .005 \). However, we are primarily concerned with values of \( p \leq .10 \) (level) and values of \( p \geq .85 \) (power) and therefore .005 is a very conservative (i.e. too large) estimate of our sampling variability. In general we feel these tables may be considered accurate to two places. All of the basic observations were standard normal variates generated using the subroutines LLRAND and NORMAL found in IRCCRAND — The Ohio State University Random Number Generator Package (see Dudewicz (1974)).
3.2. The Distribution of $\bar{F}$ (Tables of Level and Power of $\mathcal{P}_{B_1}$)

Table 3.1 presents the 10%, 5% and 1% critical points (c) of the null distribution of $\bar{F}$ needed to test (2.4.2) and the power achieved for the various combinations of $k, n_0, \delta^*$ and $z$. For example if $\delta^* = .1, n_0 = 5, k = 2$ and we want to run a 5% level test, our cutoff point $c$ is 8.00, and if we select $z = .02$ the power of our test at $\delta^* = .1$ is .311.

The critical points and powers were obtained from Program I in the following manner. For each $k$ ($k = 1, \ldots, 6$) 25 observations were drawn using the random number generator previously described. For each $\delta^*$ ($\delta^* = 0, .1, 1, 5$), $\mu_1 = \sqrt{\delta^*/2}$ was added to the observations from the first population ($k = 1$), $\mu_2 = -\sqrt{\delta^*/2}$ was added to the observations from the second population ($k = 2$) and $\mu_i = 0$ added to the observations of the remaining populations $i = 3, 4, 5, 6$ (since this is the least favorable configuration of means for the power of the $F$-test and the least favorable configuration asymptotically for $\mathcal{P}_{B_1}$ such that $\sum_{i=1}^{k} (\mu_i - \bar{u})^2 = \delta^*$ (see page 111 of Harter and Owen (1970)).

The final sample size $N_i$ for each population was determined with $z = .01$ (which yields the maximum number of observations needed since $N_i = \max(n_0 + 1, [s_i^2/z] + 1)$). The required additional observations were drawn with the appropriate $\mu_i$ added as above. Thus observations are independent between populations $i$ and $j$ ($i \neq j$) for any $\delta^*$ but are correlated within population $i$ for the various values of $\delta^*$. Then for the different combinations of $k, n_0, z$ and $\delta^*$ the $\bar{F}$ statistic.
(denoted $\tilde{F}(k,n_0,z,\delta^*)$) was calculated and compared with each value of $c = l(.25)20.75$. This process was replicated 10,000 times and $P[\tilde{F}(k,n_0,z,\delta^*) > c]$ was estimated by

$$\hat{p} = \frac{\text{No. times } \tilde{F}(k,n_0,z,\delta^*) > c}{10,000}.$$ 

The tabled values of $c$ for the 10%, 5% and 1% probabilities were obtained by setting $\delta^* = 0$, $z = 1$ and choosing the smallest $c$ such that $p \leq \alpha (\alpha = .10, .05, .01)$. Similarly to estimate the power at these levels for various values of $\delta^*$ and $z$ we used the corresponding $\hat{p}$ with $c$ determined from above. A value of 1.00 in the tables indicates that $\tilde{F}(k,n_0,z,\delta^*)$ exceeded $c$ for each of the 10,000 repetitions.

In several cases the 5% and 1% points were not obtained from the Monte Carlo calculations and intermediate values are tabled (except for $k = 6$ and $n_0 = 5$ where only the 10% point is tabled).
Table 3.1: Power of $P_{B_1}$ as a Function of $n_0$, $\alpha$ (Level) and $z$ with Associated Critical Points $c$, for $k = 2$ and $\delta^* = 0.1$

<table>
<thead>
<tr>
<th></th>
<th>$n_0 = 5$</th>
<th></th>
<th>$n_0 = 10$</th>
<th></th>
<th>$n_0 = 15$</th>
<th></th>
<th>$n_0 = 25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.00</td>
<td>8.00</td>
<td>18.50</td>
<td>5.00</td>
<td>9.50</td>
<td>3.25</td>
<td>5.00</td>
</tr>
<tr>
<td>$c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>1.01</td>
<td>.051</td>
<td>.012</td>
<td>.114</td>
<td>.055</td>
<td>.012</td>
<td></td>
</tr>
<tr>
<td>.8</td>
<td>1.04</td>
<td>.052</td>
<td>.012</td>
<td>.118</td>
<td>.057</td>
<td>.012</td>
<td></td>
</tr>
<tr>
<td>.6</td>
<td>1.07</td>
<td>.053</td>
<td>.013</td>
<td>.124</td>
<td>.060</td>
<td>.013</td>
<td></td>
</tr>
<tr>
<td>.4</td>
<td>1.12</td>
<td>.055</td>
<td>.013</td>
<td>.134</td>
<td>.065</td>
<td>.013</td>
<td></td>
</tr>
<tr>
<td>.2</td>
<td>1.31</td>
<td>.066</td>
<td>.015</td>
<td>.169</td>
<td>.085</td>
<td>.017</td>
<td></td>
</tr>
<tr>
<td>.1</td>
<td>1.71</td>
<td>.086</td>
<td>.017</td>
<td>.235</td>
<td>.127</td>
<td>.031</td>
<td></td>
</tr>
<tr>
<td>.08</td>
<td>1.94</td>
<td>.095</td>
<td>.019</td>
<td>.268</td>
<td>.154</td>
<td>.037</td>
<td></td>
</tr>
<tr>
<td>.06</td>
<td>2.29</td>
<td>1.16</td>
<td>.022</td>
<td>.319</td>
<td>.190</td>
<td>.050</td>
<td></td>
</tr>
<tr>
<td>.04</td>
<td>2.94</td>
<td>1.61</td>
<td>.028</td>
<td>.426</td>
<td>.272</td>
<td>.081</td>
<td></td>
</tr>
<tr>
<td>.02</td>
<td>5.10</td>
<td>3.11</td>
<td>.058</td>
<td>.658</td>
<td>.499</td>
<td>.214</td>
<td></td>
</tr>
<tr>
<td>.01</td>
<td>7.76</td>
<td>6.13</td>
<td>.174</td>
<td>.893</td>
<td>.800</td>
<td>.523</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.25</td>
<td>4.50</td>
<td>8.25</td>
<td>3.00</td>
<td>4.25</td>
<td>7.50</td>
<td></td>
</tr>
<tr>
<td>$c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>1.06</td>
<td>.060</td>
<td>.012</td>
<td>.112</td>
<td>.061</td>
<td>.013</td>
<td></td>
</tr>
<tr>
<td>.8</td>
<td>1.09</td>
<td>.062</td>
<td>.012</td>
<td>.116</td>
<td>.064</td>
<td>.013</td>
<td></td>
</tr>
<tr>
<td>.6</td>
<td>1.13</td>
<td>.067</td>
<td>.014</td>
<td>.121</td>
<td>.067</td>
<td>.014</td>
<td></td>
</tr>
<tr>
<td>.4</td>
<td>1.25</td>
<td>.074</td>
<td>.016</td>
<td>.134</td>
<td>.076</td>
<td>.017</td>
<td></td>
</tr>
<tr>
<td>.2</td>
<td>1.58</td>
<td>.097</td>
<td>.024</td>
<td>.168</td>
<td>.100</td>
<td>.028</td>
<td></td>
</tr>
<tr>
<td>.1</td>
<td>2.26</td>
<td>.145</td>
<td>.040</td>
<td>.238</td>
<td>.152</td>
<td>.048</td>
<td></td>
</tr>
<tr>
<td>.08</td>
<td>2.61</td>
<td>1.69</td>
<td>.048</td>
<td>.276</td>
<td>.179</td>
<td>.060</td>
<td></td>
</tr>
<tr>
<td>.06</td>
<td>3.18</td>
<td>2.15</td>
<td>.065</td>
<td>.336</td>
<td>.223</td>
<td>.080</td>
<td></td>
</tr>
<tr>
<td>.04</td>
<td>4.21</td>
<td>3.05</td>
<td>.105</td>
<td>.441</td>
<td>.321</td>
<td>.131</td>
<td></td>
</tr>
<tr>
<td>.02</td>
<td>6.57</td>
<td>5.45</td>
<td>.271</td>
<td>.684</td>
<td>.566</td>
<td>.313</td>
<td></td>
</tr>
<tr>
<td>.01</td>
<td>8.99</td>
<td>8.38</td>
<td>.606</td>
<td>.921</td>
<td>.856</td>
<td>.657</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.1 (continued): Power of $P_{E_k}$ as a Function of $n_0$, $\alpha$ (Level)
and $z$ with Associated Critical Points $c$, for $k = 2$ and $\delta^* = 1$.

<table>
<thead>
<tr>
<th>$n_0 = 5$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>5.00</td>
<td>8.00</td>
<td>18.50</td>
</tr>
<tr>
<td>$z$</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>1.0</td>
<td>.166</td>
<td>.084</td>
<td>.018</td>
</tr>
<tr>
<td>.8</td>
<td>.184</td>
<td>.095</td>
<td>.019</td>
</tr>
<tr>
<td>.6</td>
<td>.221</td>
<td>.117</td>
<td>.023</td>
</tr>
<tr>
<td>.4</td>
<td>.294</td>
<td>.155</td>
<td>.029</td>
</tr>
<tr>
<td>.2</td>
<td>.507</td>
<td>.310</td>
<td>.059</td>
</tr>
<tr>
<td>.1</td>
<td>.788</td>
<td>.618</td>
<td>.176</td>
</tr>
<tr>
<td>.08</td>
<td>.862</td>
<td>.732</td>
<td>.264</td>
</tr>
<tr>
<td>.06</td>
<td>.927</td>
<td>.849</td>
<td>.424</td>
</tr>
<tr>
<td>.04</td>
<td>.973</td>
<td>.949</td>
<td>.726</td>
</tr>
<tr>
<td>.02</td>
<td>.997</td>
<td>.994</td>
<td>.975</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 10$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.25</td>
<td>5.00</td>
<td>9.50</td>
<td></td>
</tr>
<tr>
<td>.10</td>
<td>.05</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>.239</td>
<td>.132</td>
<td>.032</td>
<td></td>
</tr>
<tr>
<td>.269</td>
<td>.153</td>
<td>.039</td>
<td></td>
</tr>
<tr>
<td>.317</td>
<td>.193</td>
<td>.053</td>
<td></td>
</tr>
<tr>
<td>.417</td>
<td>.268</td>
<td>.086</td>
<td></td>
</tr>
<tr>
<td>.655</td>
<td>.496</td>
<td>.214</td>
<td></td>
</tr>
<tr>
<td>.892</td>
<td>.803</td>
<td>.528</td>
<td></td>
</tr>
<tr>
<td>.941</td>
<td>.883</td>
<td>.664</td>
<td></td>
</tr>
<tr>
<td>.979</td>
<td>.953</td>
<td>.824</td>
<td></td>
</tr>
<tr>
<td>.996</td>
<td>.991</td>
<td>.957</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 15$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.25</td>
<td>4.50</td>
<td>8.25</td>
<td></td>
</tr>
<tr>
<td>.10</td>
<td>.05</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>.228</td>
<td>.147</td>
<td>.044</td>
<td></td>
</tr>
<tr>
<td>.263</td>
<td>.172</td>
<td>.053</td>
<td></td>
</tr>
<tr>
<td>.319</td>
<td>.216</td>
<td>.071</td>
<td></td>
</tr>
<tr>
<td>.421</td>
<td>.306</td>
<td>.111</td>
<td></td>
</tr>
<tr>
<td>.667</td>
<td>.546</td>
<td>.272</td>
<td></td>
</tr>
<tr>
<td>.907</td>
<td>.847</td>
<td>.613</td>
<td></td>
</tr>
<tr>
<td>.951</td>
<td>.933</td>
<td>.745</td>
<td></td>
</tr>
<tr>
<td>.983</td>
<td>.967</td>
<td>.882</td>
<td></td>
</tr>
<tr>
<td>.998</td>
<td>.995</td>
<td>.976</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 25$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.00</td>
<td>4.25</td>
<td>7.50</td>
<td></td>
</tr>
<tr>
<td>.10</td>
<td>.05</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>.237</td>
<td>.150</td>
<td>.049</td>
<td></td>
</tr>
<tr>
<td>.276</td>
<td>.177</td>
<td>.060</td>
<td></td>
</tr>
<tr>
<td>.331</td>
<td>.222</td>
<td>.084</td>
<td></td>
</tr>
<tr>
<td>.442</td>
<td>.314</td>
<td>.128</td>
<td></td>
</tr>
<tr>
<td>.693</td>
<td>.571</td>
<td>.307</td>
<td></td>
</tr>
<tr>
<td>.920</td>
<td>.862</td>
<td>.662</td>
<td></td>
</tr>
<tr>
<td>.957</td>
<td>.926</td>
<td>.783</td>
<td></td>
</tr>
<tr>
<td>.988</td>
<td>.972</td>
<td>.908</td>
<td></td>
</tr>
<tr>
<td>.999</td>
<td>.997</td>
<td>.984</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.1 (continued): Power of $P_{B_1}$ as a Function of $n_0$, $\alpha$ (Level) and $z$ with Associated Critical Points $c$, for $k = 2$ and $8^* = 5$.

<table>
<thead>
<tr>
<th>$n_0 = 5$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$z$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>$\alpha$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.506</td>
<td>.312</td>
<td>.058</td>
<td></td>
</tr>
<tr>
<td>.596</td>
<td>.394</td>
<td>.080</td>
<td></td>
</tr>
<tr>
<td>.712</td>
<td>.526</td>
<td>.128</td>
<td></td>
</tr>
<tr>
<td>.855</td>
<td>.728</td>
<td>.254</td>
<td></td>
</tr>
<tr>
<td>.976</td>
<td>.949</td>
<td>.724</td>
<td></td>
</tr>
<tr>
<td>.997</td>
<td>.994</td>
<td>.975</td>
<td></td>
</tr>
<tr>
<td>.998</td>
<td>.998</td>
<td>.990</td>
<td></td>
</tr>
<tr>
<td>.999</td>
<td>.999</td>
<td>.997</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 10$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$z$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>$\alpha$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.672</td>
<td>.495</td>
<td>.217</td>
<td></td>
</tr>
<tr>
<td>.761</td>
<td>.587</td>
<td>.290</td>
<td></td>
</tr>
<tr>
<td>.857</td>
<td>.722</td>
<td>.422</td>
<td></td>
</tr>
<tr>
<td>.948</td>
<td>.885</td>
<td>.655</td>
<td></td>
</tr>
<tr>
<td>.997</td>
<td>.992</td>
<td>.957</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 15$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$z$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>$\alpha$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.662</td>
<td>.545</td>
<td>.273</td>
<td></td>
</tr>
<tr>
<td>.748</td>
<td>.644</td>
<td>.366</td>
<td></td>
</tr>
<tr>
<td>.850</td>
<td>.770</td>
<td>.506</td>
<td></td>
</tr>
<tr>
<td>.950</td>
<td>.911</td>
<td>.740</td>
<td></td>
</tr>
<tr>
<td>.998</td>
<td>.996</td>
<td>.978</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 25$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$z$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>$\alpha$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.688</td>
<td>.572</td>
<td>.307</td>
<td></td>
</tr>
<tr>
<td>.771</td>
<td>.669</td>
<td>.406</td>
<td></td>
</tr>
<tr>
<td>.873</td>
<td>.788</td>
<td>.561</td>
<td></td>
</tr>
<tr>
<td>.959</td>
<td>.926</td>
<td>.779</td>
<td></td>
</tr>
<tr>
<td>.999</td>
<td>.997</td>
<td>.983</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.1 (continued): Power of $\Phi_{B_1}$ as a Function of $n_0$, $\alpha$ (Level), and $z$ with Associated Critical Points $c$, for $k = 3$ and $\delta^* = .1$

<table>
<thead>
<tr>
<th>$n_0 = 5$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>9.25</td>
<td>14.00</td>
<td>20.00</td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
<td>.025</td>
</tr>
<tr>
<td>1.</td>
<td>.101</td>
<td>.051</td>
<td>.027</td>
</tr>
<tr>
<td>.8</td>
<td>.102</td>
<td>.052</td>
<td>.027</td>
</tr>
<tr>
<td>.6</td>
<td>.104</td>
<td>.053</td>
<td>.027</td>
</tr>
<tr>
<td>.4</td>
<td>.107</td>
<td>.053</td>
<td>.028</td>
</tr>
<tr>
<td>.2</td>
<td>.115</td>
<td>.057</td>
<td>.029</td>
</tr>
<tr>
<td>.1</td>
<td>.141</td>
<td>.067</td>
<td>.034</td>
</tr>
<tr>
<td>.08</td>
<td>.149</td>
<td>.069</td>
<td>.036</td>
</tr>
<tr>
<td>.06</td>
<td>.166</td>
<td>.078</td>
<td>.038</td>
</tr>
<tr>
<td>.04</td>
<td>.209</td>
<td>.100</td>
<td>.046</td>
</tr>
<tr>
<td>.02</td>
<td>.345</td>
<td>.167</td>
<td>.077</td>
</tr>
<tr>
<td>.01</td>
<td>.623</td>
<td>.381</td>
<td>.189</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 10$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>6.00</td>
<td>8.00</td>
<td>14.00</td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>1.</td>
<td>.102</td>
<td>.053</td>
<td>.011</td>
</tr>
<tr>
<td>.8</td>
<td>.102</td>
<td>.054</td>
<td>.012</td>
</tr>
<tr>
<td>.6</td>
<td>.105</td>
<td>.055</td>
<td>.012</td>
</tr>
<tr>
<td>.4</td>
<td>.114</td>
<td>.058</td>
<td>.012</td>
</tr>
<tr>
<td>.2</td>
<td>.134</td>
<td>.071</td>
<td>.015</td>
</tr>
<tr>
<td>.1</td>
<td>.173</td>
<td>.099</td>
<td>.023</td>
</tr>
<tr>
<td>.08</td>
<td>.193</td>
<td>.111</td>
<td>.025</td>
</tr>
<tr>
<td>.06</td>
<td>.229</td>
<td>.138</td>
<td>.032</td>
</tr>
<tr>
<td>.04</td>
<td>.307</td>
<td>.194</td>
<td>.046</td>
</tr>
<tr>
<td>.02</td>
<td>.517</td>
<td>.374</td>
<td>.119</td>
</tr>
<tr>
<td>.01</td>
<td>.802</td>
<td>.086</td>
<td>.345</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 15$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>5.50</td>
<td>7.25</td>
<td>11.75</td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>1.</td>
<td>.107</td>
<td>.054</td>
<td>.011</td>
</tr>
<tr>
<td>.8</td>
<td>.108</td>
<td>.056</td>
<td>.011</td>
</tr>
<tr>
<td>.6</td>
<td>.111</td>
<td>.058</td>
<td>.012</td>
</tr>
<tr>
<td>.4</td>
<td>.118</td>
<td>.064</td>
<td>.013</td>
</tr>
<tr>
<td>.2</td>
<td>.139</td>
<td>.077</td>
<td>.018</td>
</tr>
<tr>
<td>.1</td>
<td>.186</td>
<td>.108</td>
<td>.029</td>
</tr>
<tr>
<td>.08</td>
<td>.212</td>
<td>.124</td>
<td>.035</td>
</tr>
<tr>
<td>.06</td>
<td>.254</td>
<td>.151</td>
<td>.044</td>
</tr>
<tr>
<td>.04</td>
<td>.332</td>
<td>.216</td>
<td>.067</td>
</tr>
<tr>
<td>.02</td>
<td>.555</td>
<td>.418</td>
<td>.173</td>
</tr>
<tr>
<td>.01</td>
<td>.834</td>
<td>.730</td>
<td>.461</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 25$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>5.00</td>
<td>6.75</td>
<td>10.75</td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>1.</td>
<td>.113</td>
<td>.053</td>
<td>.010</td>
</tr>
<tr>
<td>.8</td>
<td>.116</td>
<td>.055</td>
<td>.011</td>
</tr>
<tr>
<td>.6</td>
<td>.122</td>
<td>.057</td>
<td>.012</td>
</tr>
<tr>
<td>.4</td>
<td>.130</td>
<td>.063</td>
<td>.013</td>
</tr>
<tr>
<td>.2</td>
<td>.157</td>
<td>.081</td>
<td>.019</td>
</tr>
<tr>
<td>.1</td>
<td>.211</td>
<td>.119</td>
<td>.032</td>
</tr>
<tr>
<td>.08</td>
<td>.235</td>
<td>.138</td>
<td>.038</td>
</tr>
<tr>
<td>.06</td>
<td>.277</td>
<td>.170</td>
<td>.051</td>
</tr>
<tr>
<td>.04</td>
<td>.368</td>
<td>.236</td>
<td>.080</td>
</tr>
<tr>
<td>.02</td>
<td>.593</td>
<td>.443</td>
<td>.212</td>
</tr>
<tr>
<td>.01</td>
<td>.863</td>
<td>.768</td>
<td>.516</td>
</tr>
</tbody>
</table>
Table 3.1 (continued): Power of $\Phi_{B_1}$ as a Function of $n_0$, $\alpha$ (Level), and $z$ with Associated Critical Points $c$, for $k = 3$ and $\delta^* = 1$.

<table>
<thead>
<tr>
<th>$n_0 = 5$</th>
<th>$n_0 = 10$</th>
<th>$n_0 = 15$</th>
<th>$n_0 = 25$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c$</td>
<td>9.25</td>
<td>14.00</td>
<td>20.00</td>
</tr>
<tr>
<td>$\alpha$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>.133</td>
<td>.064</td>
<td>.033</td>
</tr>
<tr>
<td>.8</td>
<td>.145</td>
<td>.068</td>
<td>.035</td>
</tr>
<tr>
<td>.6</td>
<td>.165</td>
<td>.077</td>
<td>.039</td>
</tr>
<tr>
<td>.4</td>
<td>.204</td>
<td>.098</td>
<td>.047</td>
</tr>
<tr>
<td>.2</td>
<td>.3147</td>
<td>.173</td>
<td>.078</td>
</tr>
<tr>
<td>.1</td>
<td>.633</td>
<td>.391</td>
<td>.197</td>
</tr>
<tr>
<td>.08</td>
<td>.741</td>
<td>.505</td>
<td>.278</td>
</tr>
<tr>
<td>.06</td>
<td>.859</td>
<td>.677</td>
<td>.433</td>
</tr>
<tr>
<td>.04</td>
<td>.956</td>
<td>.885</td>
<td>.729</td>
</tr>
<tr>
<td>.02</td>
<td>.998</td>
<td>.992</td>
<td>.978</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
<td>1.00</td>
<td>.999</td>
</tr>
</tbody>
</table>
Table 3.1 (continued): Power of $\mathcal{P}_{B_1}$ as a Function of $n_0$, $\alpha$ (Level), and $z$ with Associated Critical Points $c$, for $k = 3$ and $\delta^* = 5$.

<table>
<thead>
<tr>
<th>$n_0 = 5$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>9.25</td>
<td>14.00</td>
<td>20.00</td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
<td>.025</td>
</tr>
<tr>
<td>1.</td>
<td>.347</td>
<td>.171</td>
<td>.079</td>
</tr>
<tr>
<td>.8</td>
<td>.425</td>
<td>.215</td>
<td>.100</td>
</tr>
<tr>
<td>.6</td>
<td>.544</td>
<td>.308</td>
<td>.149</td>
</tr>
<tr>
<td>.4</td>
<td>.742</td>
<td>.510</td>
<td>.271</td>
</tr>
<tr>
<td>.2</td>
<td>.956</td>
<td>.775</td>
<td>.726</td>
</tr>
<tr>
<td>.1</td>
<td>.998</td>
<td>.993</td>
<td>.979</td>
</tr>
<tr>
<td>.08</td>
<td>1.00</td>
<td>1.00</td>
<td>.992</td>
</tr>
<tr>
<td>.06</td>
<td>1.00</td>
<td>1.00</td>
<td>.999</td>
</tr>
<tr>
<td>.04</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.02</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 10$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6.00</td>
<td>8.00</td>
<td>14.00</td>
<td></td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>1.</td>
<td>.515</td>
<td>.369</td>
<td>.120</td>
</tr>
<tr>
<td>.8</td>
<td>.604</td>
<td>.457</td>
<td>.170</td>
</tr>
<tr>
<td>.6</td>
<td>.728</td>
<td>.594</td>
<td>.265</td>
</tr>
<tr>
<td>.4</td>
<td>.882</td>
<td>.797</td>
<td>.480</td>
</tr>
<tr>
<td>.2</td>
<td>.992</td>
<td>.983</td>
<td>.900</td>
</tr>
<tr>
<td>.1</td>
<td>1.00</td>
<td>1.00</td>
<td>.998</td>
</tr>
<tr>
<td>.08</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.06</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.04</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.02</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 15$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5.50</td>
<td>7.25</td>
<td>11.75</td>
<td></td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>1.</td>
<td>.555</td>
<td>.420</td>
<td>.182</td>
</tr>
<tr>
<td>.8</td>
<td>.650</td>
<td>.511</td>
<td>.245</td>
</tr>
<tr>
<td>.6</td>
<td>.772</td>
<td>.653</td>
<td>.670</td>
</tr>
<tr>
<td>.4</td>
<td>.910</td>
<td>.838</td>
<td>.604</td>
</tr>
<tr>
<td>.2</td>
<td>.996</td>
<td>.989</td>
<td>.947</td>
</tr>
<tr>
<td>.1</td>
<td>1.00</td>
<td>1.00</td>
<td>.999</td>
</tr>
<tr>
<td>.08</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.06</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.04</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.02</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 25$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5.00</td>
<td>6.75</td>
<td>10.75</td>
<td></td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>1.</td>
<td>.599</td>
<td>.447</td>
<td>.207</td>
</tr>
<tr>
<td>.8</td>
<td>.689</td>
<td>.544</td>
<td>.284</td>
</tr>
<tr>
<td>.6</td>
<td>.801</td>
<td>.686</td>
<td>.415</td>
</tr>
<tr>
<td>.4</td>
<td>.928</td>
<td>.864</td>
<td>.660</td>
</tr>
<tr>
<td>.2</td>
<td>.997</td>
<td>.992</td>
<td>.964</td>
</tr>
<tr>
<td>.1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.08</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.06</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.04</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.02</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 3.1 (continued): Power of $P_{B_1}$ as a Function of $n_0$, $\alpha$ (Level), and $z$ with Associated Critical Points $c$, for $k = 4$ and $b^* = .1$

<table>
<thead>
<tr>
<th>$n_0 = 5$</th>
<th>$n_0 = 10$</th>
<th>$n_0 = 15$</th>
<th>$n_0 = 25$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>13.00</td>
<td>15.25</td>
<td>18.75</td>
</tr>
<tr>
<td>$z \alpha$</td>
<td>.10</td>
<td>.075</td>
<td>.05</td>
</tr>
<tr>
<td>1.0</td>
<td>.100</td>
<td>.076</td>
<td>.051</td>
</tr>
<tr>
<td>.8</td>
<td>.101</td>
<td>.077</td>
<td>.051</td>
</tr>
<tr>
<td>.6</td>
<td>.103</td>
<td>.078</td>
<td>.052</td>
</tr>
<tr>
<td>.4</td>
<td>.105</td>
<td>.079</td>
<td>.052</td>
</tr>
<tr>
<td>.2</td>
<td>.112</td>
<td>.082</td>
<td>.056</td>
</tr>
<tr>
<td>.1</td>
<td>.124</td>
<td>.094</td>
<td>.061</td>
</tr>
<tr>
<td>.08</td>
<td>.128</td>
<td>.100</td>
<td>.066</td>
</tr>
<tr>
<td>.06</td>
<td>.142</td>
<td>.105</td>
<td>.009</td>
</tr>
<tr>
<td>.04</td>
<td>.109</td>
<td>.128</td>
<td>.083</td>
</tr>
<tr>
<td>.02</td>
<td>.271</td>
<td>.199</td>
<td>.129</td>
</tr>
<tr>
<td>.01</td>
<td>.504</td>
<td>.403</td>
<td>.286</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$c$</th>
<th>7.50</th>
<th>9.75</th>
<th>14.50</th>
<th>7.00</th>
<th>8.75</th>
<th>12.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z \alpha$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>1.0</td>
<td>.106</td>
<td>.053</td>
<td>.011</td>
<td>.105</td>
<td>.056</td>
<td>.013</td>
</tr>
<tr>
<td>.8</td>
<td>.109</td>
<td>.053</td>
<td>.012</td>
<td>.106</td>
<td>.057</td>
<td>.013</td>
</tr>
<tr>
<td>.6</td>
<td>.111</td>
<td>.055</td>
<td>.012</td>
<td>.108</td>
<td>.060</td>
<td>.013</td>
</tr>
<tr>
<td>.4</td>
<td>.117</td>
<td>.058</td>
<td>.014</td>
<td>.116</td>
<td>.064</td>
<td>.014</td>
</tr>
<tr>
<td>.2</td>
<td>.132</td>
<td>.070</td>
<td>.017</td>
<td>.136</td>
<td>.077</td>
<td>.019</td>
</tr>
<tr>
<td>.1</td>
<td>.171</td>
<td>.091</td>
<td>.025</td>
<td>.181</td>
<td>.106</td>
<td>.031</td>
</tr>
<tr>
<td>.08</td>
<td>.189</td>
<td>.103</td>
<td>.030</td>
<td>.202</td>
<td>.122</td>
<td>.037</td>
</tr>
<tr>
<td>.06</td>
<td>.222</td>
<td>.120</td>
<td>.036</td>
<td>.235</td>
<td>.148</td>
<td>.048</td>
</tr>
<tr>
<td>.04</td>
<td>.287</td>
<td>.167</td>
<td>.054</td>
<td>.307</td>
<td>.204</td>
<td>.072</td>
</tr>
<tr>
<td>.02</td>
<td>.487</td>
<td>.335</td>
<td>.135</td>
<td>.510</td>
<td>.385</td>
<td>.183</td>
</tr>
<tr>
<td>.01</td>
<td>.776</td>
<td>.649</td>
<td>.381</td>
<td>.805</td>
<td>.706</td>
<td>.461</td>
</tr>
</tbody>
</table>
Table 3.1 (continued): Power of $\mathcal{G}_{B_1}$ as a Function of $n_0$, $\alpha$ (Level), and $z$ with Associated Critical Points $c$, for $k = 4$ and $\delta^* = 1$. 

<table>
<thead>
<tr>
<th>$n_0 = 5$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>13.00</td>
<td>15.25</td>
<td>18.75</td>
</tr>
<tr>
<td>$c$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$z$</td>
<td>.10</td>
<td>.075</td>
<td>.05</td>
</tr>
<tr>
<td>.1</td>
<td>.121</td>
<td>.092</td>
<td>.061</td>
</tr>
<tr>
<td>.8</td>
<td>.129</td>
<td>.096</td>
<td>.065</td>
</tr>
<tr>
<td>.6</td>
<td>.141</td>
<td>.105</td>
<td>.070</td>
</tr>
<tr>
<td>.4</td>
<td>.169</td>
<td>.125</td>
<td>.081</td>
</tr>
<tr>
<td>.2</td>
<td>.267</td>
<td>.201</td>
<td>.131</td>
</tr>
<tr>
<td>.1</td>
<td>.518</td>
<td>.414</td>
<td>.285</td>
</tr>
<tr>
<td>.08</td>
<td>.628</td>
<td>.521</td>
<td>.379</td>
</tr>
<tr>
<td>.06</td>
<td>.773</td>
<td>.685</td>
<td>.546</td>
</tr>
<tr>
<td>.04</td>
<td>.928</td>
<td>.889</td>
<td>.806</td>
</tr>
<tr>
<td>.02</td>
<td>.996</td>
<td>.993</td>
<td>.987</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 10$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>8.00</td>
<td>10.50</td>
<td>17.50</td>
</tr>
<tr>
<td>$c$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$z$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>.1</td>
<td>.163</td>
<td>.087</td>
<td>.019</td>
</tr>
<tr>
<td>.8</td>
<td>.183</td>
<td>.099</td>
<td>.021</td>
</tr>
<tr>
<td>.6</td>
<td>.211</td>
<td>.119</td>
<td>.025</td>
</tr>
<tr>
<td>.4</td>
<td>.274</td>
<td>.162</td>
<td>.034</td>
</tr>
<tr>
<td>.2</td>
<td>.456</td>
<td>.306</td>
<td>.084</td>
</tr>
<tr>
<td>.1</td>
<td>.758</td>
<td>.614</td>
<td>.263</td>
</tr>
<tr>
<td>.08</td>
<td>.845</td>
<td>.728</td>
<td>.381</td>
</tr>
<tr>
<td>.06</td>
<td>.931</td>
<td>.862</td>
<td>.575</td>
</tr>
<tr>
<td>.04</td>
<td>.987</td>
<td>.970</td>
<td>.844</td>
</tr>
<tr>
<td>.02</td>
<td>1.00</td>
<td>1.00</td>
<td>.997</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 15$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.50</td>
<td>9.75</td>
<td>14.50</td>
</tr>
<tr>
<td>$c$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$z$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>.1</td>
<td>.170</td>
<td>.091</td>
<td>.023</td>
</tr>
<tr>
<td>.8</td>
<td>.190</td>
<td>.104</td>
<td>.028</td>
</tr>
<tr>
<td>.6</td>
<td>.226</td>
<td>.124</td>
<td>.037</td>
</tr>
<tr>
<td>.4</td>
<td>.288</td>
<td>.172</td>
<td>.057</td>
</tr>
<tr>
<td>.2</td>
<td>.490</td>
<td>.340</td>
<td>.137</td>
</tr>
<tr>
<td>.1</td>
<td>.792</td>
<td>.658</td>
<td>.385</td>
</tr>
<tr>
<td>.08</td>
<td>.874</td>
<td>.777</td>
<td>.519</td>
</tr>
<tr>
<td>.06</td>
<td>.950</td>
<td>.896</td>
<td>.717</td>
</tr>
<tr>
<td>.04</td>
<td>.992</td>
<td>.980</td>
<td>.916</td>
</tr>
<tr>
<td>.02</td>
<td>1.00</td>
<td>1.00</td>
<td>.999</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 25$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7.00</td>
<td>8.75</td>
<td>12.75</td>
</tr>
<tr>
<td>$c$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$z$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>.1</td>
<td>.174</td>
<td>.102</td>
<td>.029</td>
</tr>
<tr>
<td>.8</td>
<td>.195</td>
<td>.119</td>
<td>.037</td>
</tr>
<tr>
<td>.6</td>
<td>.229</td>
<td>.144</td>
<td>.048</td>
</tr>
<tr>
<td>.4</td>
<td>.303</td>
<td>.201</td>
<td>.074</td>
</tr>
<tr>
<td>.2</td>
<td>.517</td>
<td>.387</td>
<td>.180</td>
</tr>
<tr>
<td>.1</td>
<td>.813</td>
<td>.714</td>
<td>.468</td>
</tr>
<tr>
<td>.08</td>
<td>.891</td>
<td>.818</td>
<td>.606</td>
</tr>
<tr>
<td>.06</td>
<td>.958</td>
<td>.920</td>
<td>.784</td>
</tr>
<tr>
<td>.04</td>
<td>.996</td>
<td>.987</td>
<td>.949</td>
</tr>
<tr>
<td>.02</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 3.1 (continued): Power of $P_{B_1}$ as a Function of $n_0$, $\alpha$ (Level), and $z$ with Associated Critical Points $c$, for $k = 4$ and $\delta^* = 5$.

<table>
<thead>
<tr>
<th>$n_0 = 5$</th>
<th>$n_0 = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>13.00</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
</tr>
<tr>
<td>$z$</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>.267</td>
</tr>
<tr>
<td>.8</td>
<td>.326</td>
</tr>
<tr>
<td>.6</td>
<td>.381</td>
</tr>
<tr>
<td>.4</td>
<td>.629</td>
</tr>
<tr>
<td>.2</td>
<td>.927</td>
</tr>
<tr>
<td>.1</td>
<td>.996</td>
</tr>
<tr>
<td>.08</td>
<td>.999</td>
</tr>
<tr>
<td>.06</td>
<td>1.00</td>
</tr>
<tr>
<td>.04</td>
<td>1.00</td>
</tr>
<tr>
<td>.02</td>
<td>1.00</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 15$</th>
<th>$n_0 = 25$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>7.50</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
</tr>
<tr>
<td>$z$</td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>.488</td>
</tr>
<tr>
<td>.8</td>
<td>.577</td>
</tr>
<tr>
<td>.6</td>
<td>.706</td>
</tr>
<tr>
<td>.4</td>
<td>.866</td>
</tr>
<tr>
<td>.2</td>
<td>.993</td>
</tr>
<tr>
<td>.1</td>
<td>1.00</td>
</tr>
<tr>
<td>.08</td>
<td>1.00</td>
</tr>
<tr>
<td>.06</td>
<td>1.00</td>
</tr>
<tr>
<td>.04</td>
<td>1.00</td>
</tr>
<tr>
<td>.02</td>
<td>1.00</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 3.1 (continued): Power of $P_{\beta_1}$ as a Function of $n_0$, $\alpha$ (Level), and $z$ with Associated Critical Points $c$, for $k = 5$ and $\delta^* = .1$

<table>
<thead>
<tr>
<th>$n_0 = 5$</th>
<th>$n_0 = 10$</th>
<th>$n_0 = 15$</th>
<th>$n_0 = 25$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>$16.25$</td>
<td>$17.75$</td>
<td>$19.50$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>$10$</td>
<td>$085$</td>
<td>$07$</td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$1.0$</td>
<td>$103$</td>
<td>$085$</td>
<td>$071$</td>
</tr>
<tr>
<td>$0.8$</td>
<td>$103$</td>
<td>$086$</td>
<td>$071$</td>
</tr>
<tr>
<td>$0.6$</td>
<td>$105$</td>
<td>$087$</td>
<td>$072$</td>
</tr>
<tr>
<td>$0.4$</td>
<td>$107$</td>
<td>$090$</td>
<td>$075$</td>
</tr>
<tr>
<td>$0.2$</td>
<td>$113$</td>
<td>$094$</td>
<td>$078$</td>
</tr>
<tr>
<td>$0.1$</td>
<td>$125$</td>
<td>$106$</td>
<td>$086$</td>
</tr>
<tr>
<td>$0.08$</td>
<td>$128$</td>
<td>$109$</td>
<td>$090$</td>
</tr>
<tr>
<td>$0.06$</td>
<td>$137$</td>
<td>$113$</td>
<td>$093$</td>
</tr>
<tr>
<td>$0.04$</td>
<td>$160$</td>
<td>$133$</td>
<td>$111$</td>
</tr>
<tr>
<td>$0.02$</td>
<td>$233$</td>
<td>$196$</td>
<td>$161$</td>
</tr>
<tr>
<td>$0.01$</td>
<td>$438$</td>
<td>$380$</td>
<td>$321$</td>
</tr>
</tbody>
</table>
Table 3.1 (continued): Power of $P^{(n_0, \alpha)}$ as a Function of $n_0$, $\alpha$ (Level), and $z$ with Associated Critical Points $c$, for $k = 5$ and $\delta^* = 1$.

<table>
<thead>
<tr>
<th>$n_0 = 5$</th>
<th>$n_0 = 10$</th>
<th>$n_0 = 15$</th>
<th>$n_0 = 25$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>16.25</td>
<td>17.75</td>
<td>19.50</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
<td>.085</td>
<td>.07</td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1.8</td>
<td>.123</td>
<td>.102</td>
<td>.084</td>
</tr>
<tr>
<td>1.6</td>
<td>.129</td>
<td>.109</td>
<td>.087</td>
</tr>
<tr>
<td>1.4</td>
<td>.140</td>
<td>.115</td>
<td>.096</td>
</tr>
<tr>
<td>1.2</td>
<td>.160</td>
<td>.134</td>
<td>.110</td>
</tr>
<tr>
<td>1.0</td>
<td>.238</td>
<td>.203</td>
<td>.164</td>
</tr>
<tr>
<td>0.8</td>
<td>.443</td>
<td>.383</td>
<td>.324</td>
</tr>
<tr>
<td>0.6</td>
<td>.549</td>
<td>.486</td>
<td>.419</td>
</tr>
<tr>
<td>0.4</td>
<td>.703</td>
<td>.649</td>
<td>.579</td>
</tr>
<tr>
<td>0.2</td>
<td>.896</td>
<td>.865</td>
<td>.822</td>
</tr>
<tr>
<td>0.02</td>
<td>.995</td>
<td>.992</td>
<td>.989</td>
</tr>
<tr>
<td>0.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

|                  |             |             |             | 8.50        | 10.50       | 15.25       |
| $n_0 = 15$       | $n_0 = 25$  |             |             |             |             |             |
| $c$              | 9.50        | 11.75       | 17.25       | 10.00       | 13.00       | 20.50       |
| $\alpha$         | .10         | .05         | .01         | .10         | .05         | .01         |
| $z$              |             |             |             |             |             |             |
| 1.00             | 1.00        | 1.00        | 1.00        | 1.00        | 1.00        | 1.00        |
| 1.00             | 1.00        | 1.00        | 1.00        | .173        | .099        | .024        |
| 1.00             | 1.00        | 1.00        | 1.00        | .190        | .112        | .028        |
| 1.00             | 1.00        | 1.00        | 1.00        | .221        | .133        | .037        |
| 1.00             | 1.00        | 1.00        | 1.00        | .283        | .181        | .058        |
| 1.00             | 1.00        | 1.00        | 1.00        | .481        | .343        | .143        |
| 1.00             | 1.00        | 1.00        | 1.00        | .781        | .661        | .389        |
| 1.00             | 1.00        | 1.00        | 1.00        | .864        | .777        | .526        |
| 1.00             | 1.00        | 1.00        | 1.00        | .947        | .895        | .719        |
| 1.00             | 1.00        | 1.00        | 1.00        | .992        | .981        | .923        |
| 1.00             | 1.00        | 1.00        | 1.00        | .999        | .999        | .999        |
| 1.00             | 1.00        | 1.00        | 1.00        | 1.00        | 1.00        | 1.00        |
Table 3.1 (continued): Power of $\mathbb{P}_k$ as a Function of $n_0$, $\alpha$ (Level), and $z$ with Associated Critical Points $c$, for $k = 5$ and $\delta^* = 5$.

<table>
<thead>
<tr>
<th>$n_0 = 5$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>16.25</td>
<td>17.75</td>
<td>19.50</td>
</tr>
<tr>
<td>$z \alpha$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.239</td>
<td>0.202</td>
<td>0.165</td>
</tr>
<tr>
<td>0.8</td>
<td>0.284</td>
<td>0.243</td>
<td>0.200</td>
</tr>
<tr>
<td>0.6</td>
<td>0.367</td>
<td>0.315</td>
<td>0.264</td>
</tr>
<tr>
<td>0.4</td>
<td>0.548</td>
<td>0.488</td>
<td>0.420</td>
</tr>
<tr>
<td>0.2</td>
<td>0.895</td>
<td>0.862</td>
<td>0.823</td>
</tr>
<tr>
<td>0.1</td>
<td>0.995</td>
<td>0.992</td>
<td>0.989</td>
</tr>
<tr>
<td>0.08</td>
<td>0.999</td>
<td>0.998</td>
<td>0.997</td>
</tr>
<tr>
<td>0.06</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.04</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.02</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 10$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>10.00</td>
<td>13.00</td>
<td>20.50</td>
</tr>
<tr>
<td>$z \alpha$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.414</td>
<td>0.256</td>
<td>0.068</td>
</tr>
<tr>
<td>0.8</td>
<td>0.492</td>
<td>0.324</td>
<td>0.094</td>
</tr>
<tr>
<td>0.6</td>
<td>0.615</td>
<td>0.440</td>
<td>0.156</td>
</tr>
<tr>
<td>0.4</td>
<td>0.799</td>
<td>0.655</td>
<td>0.312</td>
</tr>
<tr>
<td>0.2</td>
<td>0.980</td>
<td>0.951</td>
<td>0.783</td>
</tr>
<tr>
<td>0.1</td>
<td>1.00</td>
<td>0.999</td>
<td>0.994</td>
</tr>
<tr>
<td>0.08</td>
<td>1.00</td>
<td>1.00</td>
<td>0.999</td>
</tr>
<tr>
<td>0.06</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.04</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.02</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 15$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>9.50</td>
<td>11.75</td>
<td>17.25</td>
</tr>
<tr>
<td>$z \alpha$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.429</td>
<td>0.299</td>
<td>0.105</td>
</tr>
<tr>
<td>0.8</td>
<td>0.512</td>
<td>0.378</td>
<td>0.148</td>
</tr>
<tr>
<td>0.6</td>
<td>0.613</td>
<td>0.502</td>
<td>0.240</td>
</tr>
<tr>
<td>0.4</td>
<td>0.825</td>
<td>0.721</td>
<td>0.440</td>
</tr>
<tr>
<td>0.2</td>
<td>0.989</td>
<td>0.971</td>
<td>0.880</td>
</tr>
<tr>
<td>0.1</td>
<td>1.00</td>
<td>1.00</td>
<td>0.999</td>
</tr>
<tr>
<td>0.08</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.06</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.04</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.02</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 25$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>8.50</td>
<td>10.50</td>
<td>15.25</td>
</tr>
<tr>
<td>$z \alpha$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.481</td>
<td>0.345</td>
<td>0.141</td>
</tr>
<tr>
<td>0.8</td>
<td>0.574</td>
<td>0.432</td>
<td>0.193</td>
</tr>
<tr>
<td>0.6</td>
<td>0.700</td>
<td>0.570</td>
<td>0.298</td>
</tr>
<tr>
<td>0.4</td>
<td>0.864</td>
<td>0.776</td>
<td>0.525</td>
</tr>
<tr>
<td>0.2</td>
<td>0.992</td>
<td>0.981</td>
<td>0.925</td>
</tr>
<tr>
<td>0.1</td>
<td>1.00</td>
<td>1.00</td>
<td>0.999</td>
</tr>
<tr>
<td>0.08</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.06</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.04</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.02</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 3.1 (continued): Power of $P_{\mu_1}$ as a Function of $n_0$, $\alpha$ (Level), and $z$ with Associated Critical Points $c$, for $k = 6$ and $\delta^* = .1$

<table>
<thead>
<tr>
<th>$n_0 = 5$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>19.75</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$z$</td>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
</tr>
<tr>
<td>1.</td>
<td>.100</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.8</td>
<td>.100</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.6</td>
<td>.101</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.4</td>
<td>.103</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.2</td>
<td>.109</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.1</td>
<td>.118</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.08</td>
<td>.123</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.06</td>
<td>.129</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.04</td>
<td>.147</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.02</td>
<td>.204</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.01</td>
<td>.378</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 10$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12.25</td>
<td>15.25</td>
<td>19.75</td>
<td></td>
</tr>
<tr>
<td>$z$</td>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
</tr>
<tr>
<td>1.</td>
<td>.104</td>
<td>.053</td>
<td>.022</td>
</tr>
<tr>
<td>.8</td>
<td>.104</td>
<td>.053</td>
<td>.022</td>
</tr>
<tr>
<td>.6</td>
<td>.105</td>
<td>.055</td>
<td>.023</td>
</tr>
<tr>
<td>.4</td>
<td>.108</td>
<td>.055</td>
<td>.023</td>
</tr>
<tr>
<td>.2</td>
<td>.117</td>
<td>.061</td>
<td>.025</td>
</tr>
<tr>
<td>.1</td>
<td>.138</td>
<td>.074</td>
<td>.031</td>
</tr>
<tr>
<td>.08</td>
<td>.148</td>
<td>.082</td>
<td>.036</td>
</tr>
<tr>
<td>.06</td>
<td>.176</td>
<td>.095</td>
<td>.041</td>
</tr>
<tr>
<td>.04</td>
<td>.218</td>
<td>.026</td>
<td>.052</td>
</tr>
<tr>
<td>.02</td>
<td>.367</td>
<td>.126</td>
<td>.052</td>
</tr>
<tr>
<td>.01</td>
<td>.638</td>
<td>.187</td>
<td>.282</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 15$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>11.25</td>
<td>13.75</td>
<td>20.25</td>
</tr>
<tr>
<td>$z$</td>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
</tr>
<tr>
<td>1.</td>
<td>.100</td>
<td>.055</td>
<td>.010</td>
</tr>
<tr>
<td>.8</td>
<td>.101</td>
<td>.055</td>
<td>.010</td>
</tr>
<tr>
<td>.6</td>
<td>.103</td>
<td>.056</td>
<td>.011</td>
</tr>
<tr>
<td>.4</td>
<td>.107</td>
<td>.059</td>
<td>.012</td>
</tr>
<tr>
<td>.2</td>
<td>.120</td>
<td>.064</td>
<td>.013</td>
</tr>
<tr>
<td>.1</td>
<td>.148</td>
<td>.079</td>
<td>.017</td>
</tr>
<tr>
<td>.08</td>
<td>.161</td>
<td>.091</td>
<td>.020</td>
</tr>
<tr>
<td>.06</td>
<td>.186</td>
<td>.104</td>
<td>.024</td>
</tr>
<tr>
<td>.04</td>
<td>.236</td>
<td>.137</td>
<td>.034</td>
</tr>
<tr>
<td>.02</td>
<td>.399</td>
<td>.263</td>
<td>.079</td>
</tr>
<tr>
<td>.01</td>
<td>.684</td>
<td>.543</td>
<td>.240</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n_0 = 25$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10.25</td>
<td>12.50</td>
<td>17.50</td>
<td></td>
</tr>
<tr>
<td>$z$</td>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
</tr>
<tr>
<td>1.</td>
<td>.107</td>
<td>.052</td>
<td>.011</td>
</tr>
<tr>
<td>.8</td>
<td>.109</td>
<td>.052</td>
<td>.012</td>
</tr>
<tr>
<td>.6</td>
<td>.110</td>
<td>.054</td>
<td>.013</td>
</tr>
<tr>
<td>.4</td>
<td>.115</td>
<td>.059</td>
<td>.014</td>
</tr>
<tr>
<td>.2</td>
<td>.131</td>
<td>.069</td>
<td>.017</td>
</tr>
<tr>
<td>.1</td>
<td>.161</td>
<td>.092</td>
<td>.023</td>
</tr>
<tr>
<td>.08</td>
<td>.177</td>
<td>.102</td>
<td>.027</td>
</tr>
<tr>
<td>.06</td>
<td>.206</td>
<td>.121</td>
<td>.033</td>
</tr>
<tr>
<td>.04</td>
<td>.264</td>
<td>.159</td>
<td>.046</td>
</tr>
<tr>
<td>.02</td>
<td>.439</td>
<td>.306</td>
<td>.122</td>
</tr>
<tr>
<td>.01</td>
<td>.731</td>
<td>.603</td>
<td>.332</td>
</tr>
</tbody>
</table>
Table 3.1 (continued): Power of $\mathcal{P}_{B_1}$ as a Function of $n_0$, $\alpha$ (Level), and $z$ with Associated Critical Points $c$, for $k = 6$ and $\delta^* = 1$. 

<table>
<thead>
<tr>
<th>$n_0 = 5$</th>
<th>$n_0 = 10$</th>
<th>$n_0 = 15$</th>
<th>$n_0 = 25$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>12.25</td>
<td>15.25</td>
<td>19.75</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.115</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.8</td>
<td>.120</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.6</td>
<td>.129</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.4</td>
<td>.146</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.2</td>
<td>.207</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.1</td>
<td>.387</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.08</td>
<td>.475</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.06</td>
<td>.630</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.04</td>
<td>.849</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.02</td>
<td>.991</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$c$</th>
<th>10.25</th>
<th>12.50</th>
<th>17.50</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>.146</td>
<td>.082</td>
<td>.018</td>
</tr>
<tr>
<td>.8</td>
<td>.159</td>
<td>.090</td>
<td>.021</td>
</tr>
<tr>
<td>.6</td>
<td>.182</td>
<td>.105</td>
<td>.026</td>
</tr>
<tr>
<td>.4</td>
<td>.234</td>
<td>.140</td>
<td>.036</td>
</tr>
<tr>
<td>.2</td>
<td>.400</td>
<td>.264</td>
<td>.082</td>
</tr>
<tr>
<td>.1</td>
<td>.692</td>
<td>.550</td>
<td>.250</td>
</tr>
<tr>
<td>.08</td>
<td>.800</td>
<td>.679</td>
<td>.354</td>
</tr>
<tr>
<td>.06</td>
<td>.904</td>
<td>.834</td>
<td>.553</td>
</tr>
<tr>
<td>.04</td>
<td>.982</td>
<td>.958</td>
<td>.826</td>
</tr>
<tr>
<td>.02</td>
<td>1.00</td>
<td>1.00</td>
<td>.997</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 3.1 (continued): Power of $\mathcal{P}_{B_1}$ as a Function of $n_0$, $\alpha$ (Level), and $z$ with Associated Critical Points $c_{_b}$, for $k = 6$ and $\theta^* = 5$.

<table>
<thead>
<tr>
<th></th>
<th>$n_0 = 5$</th>
<th></th>
<th></th>
<th>$n_0 = 10$</th>
<th></th>
<th></th>
<th>$n_0 = 15$</th>
<th></th>
<th></th>
<th>$n_0 = 25$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c$</td>
<td>19.75</td>
<td>-</td>
<td>-</td>
<td>12.25</td>
<td>-</td>
<td>-</td>
<td>10.25</td>
<td>-</td>
<td>-</td>
<td>10.25</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
<td>.10</td>
<td>.05</td>
<td>.01</td>
</tr>
<tr>
<td>$z$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>.209</td>
<td>-</td>
<td>-</td>
<td>.363</td>
<td>-</td>
<td>-</td>
<td>.435</td>
<td>-</td>
<td>-</td>
<td>.435</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.8</td>
<td>.247</td>
<td>-</td>
<td>-</td>
<td>.438</td>
<td>-</td>
<td>-</td>
<td>.523</td>
<td>-</td>
<td>-</td>
<td>.523</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.6</td>
<td>.315</td>
<td>-</td>
<td>-</td>
<td>.553</td>
<td>-</td>
<td>-</td>
<td>.654</td>
<td>-</td>
<td>-</td>
<td>.654</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.4</td>
<td>.477</td>
<td>-</td>
<td>-</td>
<td>.748</td>
<td>-</td>
<td>-</td>
<td>.831</td>
<td>-</td>
<td>-</td>
<td>.831</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.2</td>
<td>.852</td>
<td>-</td>
<td>-</td>
<td>.969</td>
<td>-</td>
<td>-</td>
<td>.988</td>
<td>-</td>
<td>-</td>
<td>.988</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.1</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.08</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.06</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.04</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.02</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.01</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
3.3. Study of the Limiting Approximations to the Distribution of $F$

In Section 2.4 we suggested approximating the distribution of $\tilde{F}$ by that of an $(n_0 - 1)/(n_0 - 3) \cdot \chi^2_{k-1}$ variate. We also noted that Hochberg (1975) suggested approximating the distribution of $\tilde{F}$ by that of $\ell \cdot F_{\ell,m}$ where $\ell$ and $m$ are selected so that $\tilde{F}$ and $\ell \cdot F_{\ell,m}$ have the same first and second moments. In this section we present Table 3.2 which compares the nominal 10%, 5% and 1% levels against the true levels for each of these approximations. In the tables the $\chi^2$ approximation is denoted by $\chi^2_{\text{TLA}}$ and Hochberg's approximation is denoted by $\chi^2_{\text{FTA}}$. These results were obtained from Program II in Appendix A in the following manner. We set $n_0 = 10$ and so 10 observations were drawn for each of six populations. The final sample size was determined and the additional observations generated. The observations from each population were multiplied by the appropriate $\sigma_k$ to obtain samples from normal populations with the correct standard deviations. $\tilde{F}$ was calculated and compared to the 10%, 5% and 1% critical points for each approximation. This process was repeated 10,000 times and the probability that $\tilde{F}$ exceeds these values was estimated by

$$\hat{\nu} = P[\tilde{F} > C_{\alpha}] = \left[ \frac{\text{No. times } \tilde{F} > C_{\alpha}}{10,000} \right],$$

where $C_{\alpha}$ represents the appropriate $\alpha$-level cutoff point. The critical points for the $\chi^2$ approximation were obtained from Owen (1962) and those for Hochberg's from Hochberg (1975). He did not give
those points for $k = 2$ or $\alpha = .01$, so these cases for his approximation are not considered here.

It is clear from Table 3.2 that these approximations are excellent and should improve for $n_0 > 10$. Neither approximation seems to dominate in any of the cases studied. Due to the simplicity of the $\chi^2$ approximation plus the control of power through the non-centrality parameter $\sum_{i=1}^{k}(\mu_i - \bar{\mu})^2/z$ offered by this approximation we suggest its use.
Table 3.2: Comparison of FTLA and FTA Approximations with True Level

<table>
<thead>
<tr>
<th>Standard Deviations $g$</th>
<th>FTLA Nominal Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>(1,1)</td>
<td>.099</td>
</tr>
<tr>
<td>(2,2)</td>
<td>.098</td>
</tr>
<tr>
<td>(1,2)</td>
<td>.100</td>
</tr>
<tr>
<td>(2,1)</td>
<td>.095</td>
</tr>
<tr>
<td>(1,3)</td>
<td>.099</td>
</tr>
<tr>
<td>(3,1)</td>
<td>.096</td>
</tr>
</tbody>
</table>
Table 3.2 (continued): Comparison of FTLA and FTA
Approximations with True Level

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
 & \text{FTLA} & & \text{FTA} & & \\
\hline
\text{Standard Deviations } & \text{Nominal Size} & 10\% & 5\% & 1\% & \text{Nominal Size} & 10\% & 5\%
\hline
\text{(1,1,1)} & .098 & .053 & .017 & & .105 & .050 & \\
\text{(1,2,3)} & .100 & .059 & .017 & & .106 & .055 & \\
\text{(3,2,1)} & .097 & .054 & .017 & & .104 & .051 & \\
\text{(1,1,3)} & .103 & .056 & .016 & & .109 & .052 & \\
\text{(3,1,1)} & .102 & .056 & .017 & & .109 & .054 & \\
\text{(2,2,2)} & .099 & .053 & .017 & & .105 & .051 & \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
 & \text{FTLA} & & \text{FTA} & & \\
\hline
\text{Standard Deviations } & \text{Nominal Size} & 10\% & 5\% & 1\% & \text{Nominal Size} & 10\% & 5\%
\hline
\text{(1,1,1,1)} & .101 & .058 & .018 & & .104 & .053 & \\
\text{(1,1,1,3)} & .107 & .058 & .018 & & .110 & .052 & \\
\text{(3,1,1,1)} & .105 & .061 & .019 & & .108 & .055 & \\
\text{(1,2,2,3)} & .103 & .057 & .019 & & .106 & .051 & \\
\text{(3,2,2,1)} & .104 & .060 & .018 & & .107 & .055 & \\
\text{(1,1,1,4)} & .107 & .058 & .018 & & .109 & .053 & \\
\hline
\end{array}
\]
Table 3.2 (continued): Comparison of FTLA and FTA Approximations with True Level

\[ k = 5 \]

<table>
<thead>
<tr>
<th>Standard Deviations ( \sigma )</th>
<th>FTLA ( \text{Nominal Size} )</th>
<th>FTA ( \text{Nominal Size} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>((1,1,1,1,1))</td>
<td>.107</td>
<td>.061</td>
</tr>
<tr>
<td>((1,1,2,3,3))</td>
<td>.105</td>
<td>.063</td>
</tr>
<tr>
<td>((1,1,4,1,4,1))</td>
<td>.106</td>
<td>.062</td>
</tr>
<tr>
<td>((3,3,2,2,1))</td>
<td>.108</td>
<td>.062</td>
</tr>
<tr>
<td>((1,1,1,1,3))</td>
<td>.109</td>
<td>.062</td>
</tr>
<tr>
<td>((3,1,1,1,1))</td>
<td>.111</td>
<td>.063</td>
</tr>
</tbody>
</table>

\[ k = 6 \]

<table>
<thead>
<tr>
<th>Standard Deviations ( \sigma )</th>
<th>FTLA ( \text{Nominal Size} )</th>
<th>FTA ( \text{Nominal Size} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>((1,1,1,1,1,1))</td>
<td>.109</td>
<td>.064</td>
</tr>
<tr>
<td>((1,1,2,2,3,3))</td>
<td>.110</td>
<td>.065</td>
</tr>
<tr>
<td>((3,3,2,2,2,1))</td>
<td>.107</td>
<td>.061</td>
</tr>
<tr>
<td>((1,1,1,1,1,4))</td>
<td>.109</td>
<td>.066</td>
</tr>
<tr>
<td>((4,1,1,1,1,1))</td>
<td>.109</td>
<td>.063</td>
</tr>
<tr>
<td>((2,2,2,2,2,2))</td>
<td>.110</td>
<td>.064</td>
</tr>
</tbody>
</table>
3.4. Non-robustness of the F-Test

We previously mentioned that several authors have claimed that the usual F-test is not robust against inequality of variances, while others seem to think it is robust. We performed a Monte Carlo sampling experiment similar to that of Brown and Forsythe (1974a) (but more extensive) to compare for \( k = 2, 3, 4, 5, 6 \) the nominal 10\%, 5\% and 1\% levels against the true levels for various configurations of \( g \) and \( n \) and to check their effect on power.

These results were obtained from Program II in the following manner. Among all the configurations of sample sizes a maximum of 20 was taken on any given population and so 20 observations were drawn for each of six populations. These observations were multiplied by the appropriate \( \sigma_i \) and \( \mu_1 = \sqrt{\sigma^*_i/2} \) was added to the observations from the first population, \( \mu_2 = -\sqrt{\sigma^*_i/2} \) was added to the observations from the second population and \( \mu_i = 0 \) for the remaining populations and the results were stored. For each value of \( k = 2, \ldots, 6 \) and each combination of \( g \) and \( n \) the F statistic was computed and compared to \( F^{\alpha}_{k, \sum_{i=1}^{g} n_i - k} \), \( \alpha = .10, .05 \) and .01. This process was repeated 10,000 times and \( P[F > F^{\alpha}_{k, \sum_{i=1}^{g} n_i - k}] \) was estimated by

\[
\hat{p} = \frac{[\text{No. times } F > F^{\alpha}_{k, \sum_{i=1}^{g} n_i - k}]}{10,000}.
\]
The results appear in Table 3.3, which gives the level, and Table 3.4, which gives the power for $\delta^* = 1.0$. As an example, from Table 3.3 if $k = 2$, $g = (2,1)$ and $n = (6,12)$ the true level of a nominal 5% test is .115. From Table 3.4 the power of this nominal 5% test at $\delta^* = 1.0$ is .490.

It should be noted that Table 3.3 is far more extensive than it first appears to be. This is true because the F-statistic is invariant under the operation of multiplication of each $\sigma_i$ by a factor of $c$. That is any entry in Table 3.3 for $g = (\sigma_1, \sigma_2, \ldots, \sigma_k)$ also gives the level for $g = (c \cdot \sigma_1, c \cdot \sigma_2, \ldots, c \cdot \sigma_k), 0 < c < \infty$.

As is well-known, the level of the F-test fluctuates significantly depending upon the configuration of $g$ and $n$. The largest deviation occurred for $k = 6$ and $n = (4,6,10,12,15,20)$. There the true level for the nominal 10% test ranged from .396 (when the smaller samples had the larger standard deviations) to .043 (when the reverse was true). Similarly, the 5% test ranged from .333 to .021 and the 1% test from .233 to .004. Other distinct deviations occurred for $k = 2,3,4$, and 5 and other configurations of $g$ and $n$. It should be noted that the deviations were less, though still sizeable, when the sample sizes for each population were equal.

Table 3.4 shows a definite fluctuation in power. It should be noted that when the sample-sizes are unequal the power depends significantly on the association (unknown to the experimenter) of standard deviations and means. For example when $k = 4$, $n = (6,6,8,10), \alpha = .10$ and $g = (1,1,1,3)$ the power is .174 while if $g = (3,1,1,1)$ the power is .345. Thus not only does the power depend upon the variances but
also upon the relationship of the $\sigma_i'$s and $\mu_i'$s. It is clear from these results that the F-test is in fact very sensitive to inequality of variances in terms of level and power and the experimenter is virtually at the mercy of the unknown variances in a single-sample experiment.

Studies were also made of the power for $\delta^*$ of .1 and .5 and similar fluctuations were noted. For example for $k = 4$, $\pi = (6,6,8,10)$ the power of the nominal 10% test ranged from .085 to .205 for $\delta^* = .1$ and from .500 to .998 for $\delta^* = .5$ over various configurations of $\pi$. 
Table 3.3: True Versus Nominal Level of the F-Test as a Function of $\eta$ and $\gamma$ for $k = 2$

<table>
<thead>
<tr>
<th>$\eta$</th>
<th>(6,6)</th>
<th>(6,12)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard Deviations $\gamma$</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10% 5% 1%</td>
</tr>
<tr>
<td>(1,1)</td>
<td>.097 .049 .010</td>
<td>.098 .048 .010</td>
</tr>
<tr>
<td>(1,2)</td>
<td>.106 .057 .012</td>
<td>.048 .020 .003</td>
</tr>
<tr>
<td>(2,1)</td>
<td>.107 .057 .013</td>
<td>.190 .115 .036</td>
</tr>
<tr>
<td>(1,3)</td>
<td>.114 .063 .014</td>
<td>.037 .016 .002</td>
</tr>
<tr>
<td>(3,1)</td>
<td>.117 .065 .017</td>
<td>.230 .153 .059</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\eta$</th>
<th>(12,12)</th>
<th>(12,20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard Deviations $\gamma$</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10% 5% 1%</td>
</tr>
<tr>
<td>(1,1)</td>
<td>.101 .050 .010</td>
<td>.104 .053 .011</td>
</tr>
<tr>
<td>(1,2)</td>
<td>.101 .053 .011</td>
<td>.059 .027 .004</td>
</tr>
<tr>
<td>(2,1)</td>
<td>.109 .052 .010</td>
<td>.166 .097 .029</td>
</tr>
<tr>
<td>(1,3)</td>
<td>.108 .056 .012</td>
<td>.051 .021 .002</td>
</tr>
<tr>
<td>(3,1)</td>
<td>.112 .057 .012</td>
<td>.194 .122 .041</td>
</tr>
</tbody>
</table>
Table 3.3 (continued): True Versus Nominal Level of the F-Test as a Function of \( \eta \) and \( \sigma \) for \( k = 3 \)

<table>
<thead>
<tr>
<th>( \eta )</th>
<th>( (6,6,6) )</th>
<th>( (6,12,15) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviations ( \sigma )</td>
<td>Nominal Size</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>(1,1,1)</td>
<td>.096</td>
<td>.048</td>
</tr>
<tr>
<td>(1,2,3)</td>
<td>.116</td>
<td>.066</td>
</tr>
<tr>
<td>(3,2,1)</td>
<td>.112</td>
<td>.067</td>
</tr>
<tr>
<td>(1,1,3)</td>
<td>.137</td>
<td>.082</td>
</tr>
<tr>
<td>(3,1,1)</td>
<td>.132</td>
<td>.085</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \eta )</th>
<th>( (11,11,11) )</th>
<th>( (8,15,20) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviations ( \sigma )</td>
<td>Nominal Size</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>(1,1,1)</td>
<td>.099</td>
<td>.051</td>
</tr>
<tr>
<td>(1,2,3)</td>
<td>.111</td>
<td>.065</td>
</tr>
<tr>
<td>(3,2,1)</td>
<td>.110</td>
<td>.063</td>
</tr>
<tr>
<td>(1,1,3)</td>
<td>.129</td>
<td>.081</td>
</tr>
<tr>
<td>(3,1,1)</td>
<td>.124</td>
<td>.075</td>
</tr>
</tbody>
</table>
Table 3.3 (continued). True Versus Nominal Level of the
F-Test as a Function of $n$ and $\sigma$
for $k = 4$

<table>
<thead>
<tr>
<th>$n$</th>
<th>(6,6,6,6)</th>
<th>(6,6,8,10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominal Size</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>Standard Deviations $\sigma$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1,1,1,1)</td>
<td>.098</td>
<td>.045</td>
</tr>
<tr>
<td>(1,1,1,3)</td>
<td>.140</td>
<td>.092</td>
</tr>
<tr>
<td>(3,1,1,1)</td>
<td>.141</td>
<td>.093</td>
</tr>
<tr>
<td>(1,2,2,3)</td>
<td>.113</td>
<td>.064</td>
</tr>
<tr>
<td>(3,2,2,1)</td>
<td>.111</td>
<td>.065</td>
</tr>
<tr>
<td>(1,1,1,4)</td>
<td>.156</td>
<td>.107</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>(12,12,12,12)</th>
<th>(6,10,16,20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominal Size</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>Standard Deviations $\sigma$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1,1,1,1)</td>
<td>.097</td>
<td>.049</td>
</tr>
<tr>
<td>(1,1,1,3)</td>
<td>.132</td>
<td>.089</td>
</tr>
<tr>
<td>(3,1,1,1)</td>
<td>.134</td>
<td>.087</td>
</tr>
<tr>
<td>(1,2,2,3)</td>
<td>.113</td>
<td>.063</td>
</tr>
<tr>
<td>(3,2,2,1)</td>
<td>.111</td>
<td>.062</td>
</tr>
<tr>
<td>(1,1,1,4)</td>
<td>.144</td>
<td>.101</td>
</tr>
</tbody>
</table>
Table 3.3 (continued): True Versus Nominal Level of the F-Test as a Function of $\eta$ and $\gamma$

for $k = 5$

<table>
<thead>
<tr>
<th>$\eta$</th>
<th>(6,6,6,6)</th>
<th>(6,8,8,12,12)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominal Size</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
</tr>
<tr>
<td>Standard Deviations $\gamma$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1,1,1,1,1)</td>
<td>.097 .046 .009</td>
<td>.104 .049 .008</td>
</tr>
<tr>
<td>(1,1,2,3,3)</td>
<td>.127 .072 .025</td>
<td>.071 .039 .010</td>
</tr>
<tr>
<td>(1,4,1,4,1)</td>
<td>.144 .093 .035</td>
<td>.113 .070 .024</td>
</tr>
<tr>
<td>(3,3,2,2,1)</td>
<td>.113 .061 .018</td>
<td>.169 .104 .031</td>
</tr>
<tr>
<td>(1,1,1,1,3)</td>
<td>.147 .098 .042</td>
<td>.095 .059 .022</td>
</tr>
<tr>
<td>(3,1,1,1,1)</td>
<td>.146 .101 .044</td>
<td>.227 .168 .089</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\eta$</th>
<th>(13,13,13,13,13)</th>
<th>(6,10,16,18,20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominal Size</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
</tr>
<tr>
<td>Standard Deviations $\gamma$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1,1,1,1,1)</td>
<td>.098 .047 .008</td>
<td>.096 .048 .010</td>
</tr>
<tr>
<td>(1,1,2,2,3)</td>
<td>.122 .073 .021</td>
<td>.055 .029 .008</td>
</tr>
<tr>
<td>(1,4,1,4,1)</td>
<td>.136 .086 .030</td>
<td>.135 .085 .032</td>
</tr>
<tr>
<td>(3,3,2,2,1)</td>
<td>.110 .062 .016</td>
<td>.215 .137 .050</td>
</tr>
<tr>
<td>(1,1,1,1,3)</td>
<td>.138 .095 .043</td>
<td>.068 .044 .016</td>
</tr>
<tr>
<td>(3,1,1,1,1)</td>
<td>.141 .091 .038</td>
<td>.296 .231 .135</td>
</tr>
</tbody>
</table>
Table 3.3 (continued): True Versus Nominal Level of the F-Test as a Function of \( n \) and \( \sigma \)
for \( k = 6 \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>((6,6,6,6,6,6))</th>
<th>((4,6,10,12,15,20))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviations ( \sigma )</td>
<td>Nominal Size</td>
<td>10%</td>
</tr>
<tr>
<td>((1,1,1,1,1,1))</td>
<td>.098</td>
<td>.046</td>
</tr>
<tr>
<td>((1,1,2,2,3,3))</td>
<td>.124</td>
<td>.073</td>
</tr>
<tr>
<td>((3,3,2,2,1,1))</td>
<td>.123</td>
<td>.070</td>
</tr>
<tr>
<td>((1,1,1,1,1,4))</td>
<td>.173</td>
<td>.130</td>
</tr>
<tr>
<td>((4,1,1,1,1,1))</td>
<td>.167</td>
<td>.127</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( n )</th>
<th>((11,11,11,11,11,11))</th>
<th>((6,6,10,10,20,20))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviations ( \sigma )</td>
<td>Nominal Size</td>
<td>10%</td>
</tr>
<tr>
<td>((1,1,1,1,1,1))</td>
<td>.101</td>
<td>.053</td>
</tr>
<tr>
<td>((1,1,2,2,3,3))</td>
<td>.127</td>
<td>.075</td>
</tr>
<tr>
<td>((3,3,2,2,1,1))</td>
<td>.123</td>
<td>.071</td>
</tr>
<tr>
<td>((1,1,1,1,1,4))</td>
<td>.158</td>
<td>.116</td>
</tr>
<tr>
<td>((4,1,1,1,1,1))</td>
<td>.157</td>
<td>.115</td>
</tr>
<tr>
<td>n</td>
<td>$(6,6)$</td>
<td>$(6,12)$</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Standard Deviations $\sigma$</td>
<td>Nominal Size</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>$(1,1)$</td>
<td>.734</td>
<td>.598</td>
</tr>
<tr>
<td>$(2,2)$</td>
<td>.304</td>
<td>.193</td>
</tr>
<tr>
<td>$(1,2)$</td>
<td>.419</td>
<td>.288</td>
</tr>
<tr>
<td>$(2,1)$</td>
<td>.427</td>
<td>.298</td>
</tr>
<tr>
<td>$(1,3)$</td>
<td>.278</td>
<td>.184</td>
</tr>
<tr>
<td>$(3,1)$</td>
<td>.289</td>
<td>.189</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n</th>
<th>$(12,12)$</th>
<th>$(12,20)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviations $\sigma$</td>
<td>Nominal Size</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>$(1,1)$</td>
<td>.958</td>
<td>.912</td>
</tr>
<tr>
<td>$(2,2)$</td>
<td>.511</td>
<td>.381</td>
</tr>
<tr>
<td>$(1,2)$</td>
<td>.687</td>
<td>.554</td>
</tr>
<tr>
<td>$(2,1)$</td>
<td>.680</td>
<td>.553</td>
</tr>
<tr>
<td>$(1,3)$</td>
<td>.448</td>
<td>.325</td>
</tr>
<tr>
<td>$(3,1)$</td>
<td>.449</td>
<td>.327</td>
</tr>
</tbody>
</table>
Table 3.4 (continued): Power of the F-Test as a Function of \( \tau, \eta \) and the Nominal Size for \( k = 3 \) and \( S^* = 1.0 \)

<table>
<thead>
<tr>
<th>( \eta )</th>
<th>( (6,6,6) )</th>
<th>( (6,12,15) )</th>
<th>( (11,11,11) )</th>
<th>( (8,15,20) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviations ( \sigma )</td>
<td>Nominal Size</td>
<td></td>
<td>Nominal Size</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
</tr>
<tr>
<td>(1,1,1)</td>
<td>.639</td>
<td>.488</td>
<td>.226</td>
<td>.809</td>
</tr>
<tr>
<td>(1,2,3)</td>
<td>.235</td>
<td>.141</td>
<td>.041</td>
<td>.181</td>
</tr>
<tr>
<td>(3,2,1)</td>
<td>.234</td>
<td>.155</td>
<td>.055</td>
<td>.422</td>
</tr>
<tr>
<td>(1,1,3)</td>
<td>.286</td>
<td>.178</td>
<td>.059</td>
<td>.213</td>
</tr>
<tr>
<td>(3,1,1)</td>
<td>.276</td>
<td>.190</td>
<td>.082</td>
<td>.528</td>
</tr>
<tr>
<td>(2,2,2)</td>
<td>.264</td>
<td>.149</td>
<td>.042</td>
<td>.329</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1,1,1)</td>
<td>.895</td>
<td>.816</td>
<td>.580</td>
<td>.896</td>
</tr>
<tr>
<td>(1,2,3)</td>
<td>.366</td>
<td>.235</td>
<td>.077</td>
<td>.229</td>
</tr>
<tr>
<td>(3,2,1)</td>
<td>.340</td>
<td>.246</td>
<td>.110</td>
<td>.480</td>
</tr>
<tr>
<td>(1,1,3)</td>
<td>.460</td>
<td>.301</td>
<td>.104</td>
<td>.266</td>
</tr>
<tr>
<td>(3,1,1)</td>
<td>.391</td>
<td>.297</td>
<td>.146</td>
<td>.585</td>
</tr>
<tr>
<td>(2,2,2)</td>
<td>.392</td>
<td>.271</td>
<td>.104</td>
<td>.402</td>
</tr>
</tbody>
</table>
Table 3.4 (continued): Power of the F-Test as a Function of $g$, $n$ and the Nominal Size for $k = 4$ and $s^* = 1.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>(6,6,6,6)</th>
<th>(6,6,8,10)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominal Size</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>Standard Deviations $g$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1,1,1,1)</td>
<td>.572</td>
<td>.128</td>
</tr>
<tr>
<td>(1,1,1,3)</td>
<td>.285</td>
<td>.187</td>
</tr>
<tr>
<td>(3,1,1,1)</td>
<td>.275</td>
<td>.197</td>
</tr>
<tr>
<td>(1,2,2,3)</td>
<td>.206</td>
<td>.123</td>
</tr>
<tr>
<td>(3,2,2,1)</td>
<td>.218</td>
<td>.142</td>
</tr>
<tr>
<td>(1,1,1,4)</td>
<td>.234</td>
<td>.160</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>12,12,12,12</th>
<th>6,10,16,20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominal Size</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>Standard Deviations $g$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1,1,1,1)</td>
<td>.884</td>
<td>.800</td>
</tr>
<tr>
<td>(1,1,1,3)</td>
<td>.480</td>
<td>.332</td>
</tr>
<tr>
<td>(3,1,1,1)</td>
<td>.410</td>
<td>.317</td>
</tr>
<tr>
<td>(1,2,2,3)</td>
<td>.342</td>
<td>.214</td>
</tr>
<tr>
<td>(3,2,2,1)</td>
<td>.331</td>
<td>.234</td>
</tr>
<tr>
<td>(1,1,1,4)</td>
<td>.325</td>
<td>.215</td>
</tr>
</tbody>
</table>
Table 3.4 (continued): Power of the F-Test as a Function of $\varphi$, $\eta$ and the Nominal Size for $k = 5$ and $\delta^* = 1.0$

<table>
<thead>
<tr>
<th>$n$</th>
<th>(6,6,6,6)</th>
<th>(6,8,8,12,12)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard Deviations $\varphi$</td>
<td>Nominal Size 10%</td>
</tr>
<tr>
<td></td>
<td>(1,1,1,1,1)</td>
<td>.529</td>
</tr>
<tr>
<td></td>
<td>(1,1,2,3,3)</td>
<td>.188</td>
</tr>
<tr>
<td></td>
<td>(1,4,1,4,1)</td>
<td>.187</td>
</tr>
<tr>
<td></td>
<td>(3,3,2,2,1)</td>
<td>.188</td>
</tr>
<tr>
<td></td>
<td>(1,1,1,1,3)</td>
<td>.285</td>
</tr>
<tr>
<td></td>
<td>(3,1,1,1,1)</td>
<td>.276</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n$</th>
<th>(13,13,13,13)</th>
<th>(6,10,16,18,20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard Deviations $\varphi$</td>
<td>Nominal Size 10%</td>
</tr>
<tr>
<td></td>
<td>(1,1,1,1,1)</td>
<td>.878</td>
</tr>
<tr>
<td></td>
<td>(1,1,2,3,3)</td>
<td>.295</td>
</tr>
<tr>
<td></td>
<td>(1,4,1,4,1)</td>
<td>.253</td>
</tr>
<tr>
<td></td>
<td>(3,3,2,2,1)</td>
<td>.288</td>
</tr>
<tr>
<td></td>
<td>(1,1,1,1,3)</td>
<td>.494</td>
</tr>
<tr>
<td></td>
<td>(3,1,1,1,1)</td>
<td>.425</td>
</tr>
</tbody>
</table>
Table 3.4 (continued): Power of the F-Test as a Function of $\gamma$, $n$, and the Nominal Size for $k = 6$ and $\delta^* = 1.0$

<table>
<thead>
<tr>
<th>$\eta$</th>
<th>$n = (6, 6, 6, 6, 6)$</th>
<th>$n = (4, 6, 10, 12, 15, 20)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard Deviations $\gamma$</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$(1,1,1,1,1,1)$</td>
<td>.295</td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$(1,1,2,2,3,3)$</td>
<td>.188</td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$(3,3,2,2,1,1)$</td>
<td>.108</td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$(1,1,1,1,1,4)$</td>
<td>.257</td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$(4,1,1,1,1,1)$</td>
<td>.246</td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$(2,2,2,2,2)$</td>
<td>.209</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\eta^*$</th>
<th>$n = (11, 1, 1, 1, 1, 1)$</th>
<th>$n = (6, 6, 10, 10, 20, 20)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard Deviations $\gamma$</td>
<td>Nominal Size</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$(1,1,1,1,1,1)$</td>
<td>.792</td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$(1,1,2,2,3,3)$</td>
<td>.258</td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$(3,3,2,2,1,1)$</td>
<td>.268</td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$(1,1,1,1,1,4)$</td>
<td>.325</td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$(4,1,1,1,1,1)$</td>
<td>.305</td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$(2,2,2,2,2)$</td>
<td>.296</td>
</tr>
</tbody>
</table>
3.5. Relative Efficiency of $F$ and $\tilde{F}$

Based upon the fact that the properties of $\mathcal{F}_{B_1}$ are independent of the unknown variances and the results of Section 3.4, it appears that the better choice of test procedures are those based on $\mathcal{F}_{B_1}$. However, when using $\mathcal{F}_{B_1}$ the sample sizes are random variables and one would like to compare the sample sizes of the two procedures.

We define the relative efficiency of the $F$-test and $\mathcal{F}_{B_1}$ by

$$R(\alpha, \beta, \delta^*) = \frac{E(N)}{\sum_{i=1}^{k} n_i}$$

(3.5.1)

where $\sum_{i=1}^{k} n_i$ is the total sample size required by the $F$-test and $E(N)$ is the expected total sample size required by $\mathcal{F}_{B_1}$ to obtain power $\beta$ for level $\alpha$ at the alternative $\delta^*$. We estimate this relative efficiency by

$$R = \frac{\sum_{i=1}^{k} \sigma_i^2 / z}{\sum_{i=1}^{k} n_i}$$

since $E(N) \approx \sum_{i=1}^{k} \sigma_i^2 / z$ as long as $n_0$ is not close to $\sigma_i^2 / z$. Using the results of Tables 3.1, 3.3 and 3.4 we are able to give some idea of the range of values of $R$ for various values of $\alpha$, $\beta$ and $\sigma$ with $\delta^* = 1.0$. Table 3.5 contains a lower and upper bound for $R$ obtained by fixing $\gamma$ and using Table 3.3 to find $\alpha$ and Table 3.4 to find $\beta$ for the particular values of $\sigma$ for the $F$-test. Then for that $\alpha$ Table 3.1 was used to obtain $z_1$ and $z_2$ such that $z_1 < z_2$ and the power using $\mathcal{F}_{B_1}$ with $z_1$ is less than $\beta$ but
the power using $\mathcal{P}_{B_1}$ with $z_2$ is greater than $\beta$. Thus the actual $z$ needed to achieve the same power would satisfy $z_1 < z < z_2$.

TABLE 3.5

<table>
<thead>
<tr>
<th>No. of Populations</th>
<th>Deviations</th>
<th>Sample Sizes</th>
<th>$R(\alpha, \beta, 1.0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>(3,1)</td>
<td>(12,20)</td>
<td>$R(0.041,0.287,1.) = .781(1.04)$</td>
</tr>
<tr>
<td>4</td>
<td>(1,1,1,1)</td>
<td>(6,10,16,20)</td>
<td>$R(0.050,0.607,1.) = .385(0.769)$</td>
</tr>
<tr>
<td>4</td>
<td>(1,2,2,3)</td>
<td>(6,10,16,20)</td>
<td>$R(0.049,0.138,1.) = .500(0.565)$</td>
</tr>
<tr>
<td>5</td>
<td>(1,1,1,1,1)</td>
<td>(6,10,16,18,20)</td>
<td>$R(0.050,0.563,1.) = .714(0.893)$</td>
</tr>
<tr>
<td>5</td>
<td>(3,3,2,2,1)</td>
<td>(6,10,16,18,20)</td>
<td>$R(0.137,0.252,1.) = .643(0.771)$</td>
</tr>
<tr>
<td>6</td>
<td>(1,1,1,1,1,1)</td>
<td>(6,6,10,10,20,20)</td>
<td>$R(0.050,0.394,1.) = .417(0.833)$</td>
</tr>
<tr>
<td>6</td>
<td>(1,1,2,2,3,3)</td>
<td>(6,6,10,10,20,20)</td>
<td>$R(0.039,0.057,1.) = .354(0.389)$</td>
</tr>
</tbody>
</table>

It should be noted that because $n_0 = 10$ the case of $\bar{y} = (12,12)$ for $k = 2$, $\bar{y} = (12,12,12,12)$ for $k = 4$ (13,13,13,13,13) for $k = 5$ are not really reflective of the true situation since essentially $\mathcal{P}_{B_1}$ starts out taking more observations than really required to achieve the same power as the F-test. Also unequal sample sizes prove to be disastrous to the F-test whenever the variances are unequal.
These results are by no means complete and additional research is planned with regard to this relative efficiency. However, the results do indicate that in general the two-stage procedure is significantly more efficient than the single-stage procedure with unequal sample sizes. Even in the case of equal variances and equal sample sizes the two-stage procedure is not performing too badly especially in light of the fact that we were conservative in our choice of $n_0$ and of $z_1$ and $z_2$ since we took the level of $\mathcal{P}_{B_1}$ to be the largest level $\leq \alpha$ available from Table 3.1. For example if $n = (13, 13, 13, 13)$ and $g = (1, 1, 1, 1)$, $R(.05, .793, 1.) = .961(1.28)$. 
CHAPTER IV

HETEROSCEDASTIC DECISION THEORY
AND THE HETEROSCEDASTIC METHOD

4.1. The Statistical Decision Problem

In this chapter we consider general statistical decision problems in which observations are from \( k \) independent \( p \)-dimensional multivariate normal populations with unknown mean vectors and unknown covariance matrices. Our goal is to develop (sampling and decision) rules for which the risk function is both independent of the unknown covariance matrices and has specified properties. The estimation and testing problems of Chapter 2 may be derived from the general theory given here.

The general decision problem is specified as follows. Let \( \pi_1, \pi_2, \ldots, \pi_k \) be \( k \) independent \( p \)-dimensional multivariate normal populations such that \( \pi_i \sim \mathcal{N}_p(\mu_i, \Sigma_i) \), where \( \mu_i = (\mu_{i1}, \ldots, \mu_{ip}) \) and \( \Sigma_i = (\sigma_{ij}) \).

Let \( \Theta = \Theta_1 \times \Theta_2 \) where \( \Theta_1 = \{ \mu = (\mu_1, \ldots, \mu_k): \mu_i \in \mathbb{R}^p \} \) and \( \Theta_2 = \{ \Sigma = (\Sigma_1, \ldots, \Sigma_k): \Sigma_i \) is a positive-definite symmetric matrix of order \( p \} \). Let \( A \) be a nonempty set of actions, let \( \ell(\theta, a) \) be a real-valued loss function defined on \( \Theta \times A \), and let
\( \hat{X} \) denote the vector of observations \( \hat{X} = (X_1, X_2, \ldots, X_k) \) and \( \hat{X}_1 = (X_{11}, X_{12}, \ldots) \). (Since we will consider sequential, in particular, two-stage procedures, we regard the observation vector \( \hat{X}_1 \) for \( \pi_i \) as having possibly infinite dimension.) Thus in general we consider the statistical decision problem

\[
(\Theta, \mathcal{A}, \ell), \hat{X}.
\] (4.1.1)

We prove in Section 4.4 that (under certain conditions) no single-stage procedure exists which satisfies our goals for this decision problem. We now consider procedures with two-stage sampling which yield a risk function both independent of the unknown covariance matrices and having specified properties relating to \( u_1, \ldots, u_k \) (but independent of \( t_1, \ldots, t_k \)).
4.2. The Two-Stage Sampling Procedure

Our sampling for each population will follow the two-stage scheme of Chatterjee (1959a) discussed in Section 1.3. Let \( X_{ir} \) correspond to (1.3.13) for the \( i \)th population \((i = 1, 2, \ldots, k, r = 1, 2, \ldots, p)\) and construct the \( p \)-dimensional vector \( \hat{X}_i = (X_{i1}, \ldots, X_{ip})' \).

It follows from page 125 of Chatterjee (1959a) that the conditional distribution of \( \hat{X}_i \) given \( S_i = (s_{ilm}) \) (the sample covariance matrix for \( \pi_i \) based on the first \( n_0 \) observations) is \( N_p(\mu_i, \Sigma_i(S_i)) \) where

\[
\Sigma_i(S_i) = (z\alpha^{rs} \sum_{l=1}^{p} \sum_{m=1}^{p} \sigma_{ilm} s_{ilm}). \tag{4.2.1}
\]

Letting \( L_i = \sum_{l=1}^{p} \sum_{m=1}^{p} \sigma_{ilm} \) we may write the conditional covariance matrix as \( \Sigma_i(S_i) = zL_i(\alpha^{rs}) \), and the conditional density of \( \hat{X}_i \) may be expressed as

\[
f(x_{i1}, x_{i2}, \ldots, x_{ip} | S_i) = \frac{1}{(2\pi)^{p/2} |zL_i(\alpha^{rs})|} \exp\left\{ -\frac{1}{2} (x_i - \mu_i)' \left[ \frac{(\alpha^{rs})}{zL_i} \right] (x_i - \mu_i) \right\} \tag{4.2.2}
\]

where \( x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})' \). If we let \( V_i = (n_0-1)S_i \) and denote \( V_i = (V_{ilm}) \) then we may replace \( L_i \) by \((n_0-1)L_i'\) where \( L_i' = \sum_{l=1}^{p} \sum_{m=1}^{p} \sigma_{ilm} V_{ilm} \). Now \( L_i' \) is the trace of the matrix
$(\sigma_{i\ell m})(y^i\ell m)$ and hence may be expressed as the sum of the roots of the determinantal equation

$$|(\sigma_{i\ell m})(y^i\ell m) - \lambda I| = 0$$

\hspace{1cm} (4.2.3)

or

$$|(\sigma_{i\ell m}) - \lambda(y^i\ell m)| = 0.$$ \hspace{1cm} (4.2.4)

The roots of this equation are reciprocals of the roots of the equation

$$|(y^i\ell m) - \lambda(\sigma_{i\ell m})| = 0.$$ \hspace{1cm} (4.2.5)

Now $(\sigma_{i\ell m})$ is positive-definite and $(y^i\ell m)$ is positive-definite with probability one, therefore the roots of (4.2.5) are all real and positive, say $\lambda_1 < \lambda_{i_2} < \ldots < \lambda_{i_l}$. Replacing $L_1$ by

$$(n_0 - 1) \sum_{\ell=1}^{p} 1/\lambda_{i\ell}$$

in (4.2.2) we see that the conditional joint density of $\tilde{x}_i$ depends upon $S_i$ only through $L_1^i$. Now the joint density of $\lambda_{i1}, \ldots, \lambda_{ip}$ is given by $p(\lambda_{i1}, \ldots, \lambda_{ip})$ defined by (1.3.21), and we may express the unconditioned joint density function of $\tilde{x}_i$ by

$$h_i(x_{i1}, \ldots, x_{ip})$$

\hspace{1cm} (4.2.6)

$$= \int_{0 < \lambda_{i1} < \cdots < \lambda_{i1}} \frac{1}{(2\pi)^{p/2} |z(n_0 - 1) \sum_{\ell=1}^{p} 1/\lambda_{i\ell}(\alpha^{rs})|}
\times \exp \left\{ - \frac{1}{2} (x - \mu_i)' \left[ \frac{1}{z^2 \sum_{\ell=1}^{p} 1/\lambda_{i\ell}} \right] (x - \mu_i) \right\}
\times p(\lambda_{i1}, \ldots, \lambda_{ip}) \, d\lambda_{i1}, \ldots, d\lambda_{ip},$$

\hspace{1cm} (4.2.6)
which is independent of \( \hat{\mu} \).

Now \( \tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_k \) are independent vectors hence the joint density of \( \tilde{X} = (\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_k) \) is

\[
f(\tilde{X}|\mu) = \prod_{i=1}^{k} h_i(x_{i1}, \ldots, x_{ip}) = \prod_{i=1}^{k} h_i(x_{i1}, \ldots, x_{ip})
\]  \( (4.2.7) \)

where \( \tilde{X} = (x_{11}, \ldots, x_{1p}, x_{21}, \ldots, x_{2p}, \ldots, x_{kp}) \).

We denote the joint distribution function of \( \tilde{X}_i \) by

\[
F_i(x_i|\mu_i) = \int_{-\infty}^{x_{il}} \cdots \int_{-\infty}^{x_{ip}} h_i(v_{i1}, \ldots, v_{ip}) dv_{i1} \cdots dv_{ip}
\]  \( (4.2.8) \)

and the joint distribution function of \( \tilde{X} \) by

\[
F(\tilde{X}|\mu) = \prod_{i=1}^{k} F_i(x_i|\mu_i)
\]  \( (4.2.9) \)

We are then faced with the following statistical decision problem

\[
(\Theta, A, \delta), \tilde{X} \quad \text{with} \quad \tilde{X} \sim F(\tilde{X}|\mu).
\]  \( (4.2.10) \)

Also note that if we have \( k = 1, p = 1 \) and \( (\alpha_{rs}) = 1 \), then the above reduces to Stein's two-stage sampling scheme \( \mathcal{D}_S \) for one univariate normal population and \( \tilde{X} \) reduces to the generalized sample mean.
It can now be shown that, for the above statistical decision problem, the risk function associated with any decision rule (randomized or nonrandomized) will be independent of $\mu_1, \ldots, \mu_k$.

Theorem (4.2.11). The risk function $R(\mu, d)$ defined for statistical decision problem (4.2.10) and the decision rule $d$ (randomized or nonrandomized) is independent of the covariance matrices.

Proof. If $d$ is a nonrandomized decision rule then

$$R(\mu, d) = \int \ell(\mu, d(x)) \, dF(x|\mu).$$

Since $F(x|\mu)$ is independent of $\mu_1, \ldots, \mu_k$, so is $R(\mu, d)$. If $d$ is randomized then $d(x)$ is a probability distribution $P_x$ over $A$. In this case

$$R(\mu, d) = \int \int \ell(\mu, v) \, dP_x(v) \, dF(x|\mu)$$

and again $P_x(v)$ and $F(x|\mu)$ are independent of $\mu_1, \ldots, \mu_k$, hence so is $R(\mu, d)$.

It is clear that the two-stage sampling scheme and the generalized mean vectors lead to risk functions which are independent of the unknown covariance matrices. It is not clear, however, which decision rule should be chosen. We propose the following.
The Heteroscedastic Method (4.2.12). In order to decide which decision rule \( d \) to use for the decision problem \((\Theta, A, \ell), \mathfrak{X}\) consider the decision rule you would use for the corresponding decision problem \((\Theta, A, \ell), \mathfrak{Y}\) with \( \mathfrak{Y} = (\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k)' \) where \( \{\overline{X}_i\}_{i=1}^k \) are independent, \( \overline{X}_i \) being the mean vector of a sample based on \( N \) observation from \( \pi_i \sim N(\mu_i, \ell) \), \( \ell \) known \( i = 1, 2, \ldots, k \). That is, consider the same decision problem but with observations from single-stage samples of size \( N \) from \( k \) multivariate normal populations with mean \( \mu_i \) and common known covariance matrix \( \ell \).

If in this case decision rule \( d \) (in general a function of \( \overline{X}_1, \ldots, \overline{X}_k \) and \( \ell/N \)) would be used, use the same rule \( d \) for the problem \((\Theta, A, \ell), \mathfrak{X}\) after replacing \( \overline{X}_i \) with \( \overline{X}_i \) and \( \ell/N \) with \( z(\alpha^{RS}) \).

The intuitive justification for this choice of decision rule is as follows. When making inferences about the mean vectors of multivariate normal distributions the most favorable situation for the statistician is to have the covariance matrices known and equal, since in such a situation there are no nuisance parameters. The two-stage sampling scheme is designed to choose the final sample-size in such a way as to enable the data to appear as if they come from normal populations with equal known covariance matrices. The common covariance matrix is given by \( (\alpha^{RS}) \), which is chosen by the experimenter. \( z(\alpha^{RS}) \) in general plays the role of \( (1/N)^{1/2} \). This correspondence is most easily seen when \( p = 1 \) and \( k = 1 \). In this case if we are interested in testing that the mean \( \mu \) is equal to \( \mu_0 \) and
\( \sigma^2 \) is known, we would take a sample of size \( N \) and base our test on the statistic \( \sqrt{N}(\bar{X} - \mu) / \sigma \) where \( \bar{X} \) is the sample mean. When \( \sigma^2 \) is unknown the two-stage procedure yields a test based on \( \bar{X} - \mu_0 / \sqrt{z} \). Thus \( \bar{X} \) plays the role of \( \bar{X} \) and \( z \) plays the role of \( \sigma^2 / N \). Thus the two-stage procedure forces the problem to act like a single-stage problem when the covariance matrices are known and equal.

An important property of the Heteroscedastic Method is the following. Let \( R_N(\mu, d) \) denote the risk function associated with rule \( d \) in the case of \( \mu \) known and equal for each population and \( N \) is the common sample size. Let \( \lim_{N \to \infty} R_N(\mu, d) = \ell(\mu, d(\mu)) \) where \( \ell \) is the loss function associated with \( R_N(\mu, d) \). Further let \( R_z(\mu, d) \) denote the risk function for the corresponding two-stage procedure with parameter \( z \).

**Corollary (4.2.13).** The distribution of \( \bar{X} \), as \( z \to 0 \), is degenerate at \( \mu \).

**Proof.** \( \lim_{z \to 0} F(\bar{X} | \mu) = \lim_{z \to 0} \int \Pi_{i=1}^k F_i(\bar{X}_i | S_i) \, dP_i(S_i) \) where \( F_i(\bar{X}_i | S_i) \) is the conditional distribution function of \( \bar{X}_i \) given \( S_i \) and \( P_i(S_i) \) is the distribution function of \( S_i \). Now \( \Pi_{i=1}^k F_i(\bar{X}_i | S_i) \) is bounded and continuous hence the limit may be brought under the integral sign so \( \lim_{z \to 0} F(\bar{X} | \mu) \)

\[ = \int \lim_{z \to 0} \Pi_{i=1}^k F_i(\bar{X}_i | S_i) \, dP_i(S_i) \]  

The conditional distribution of \( \bar{X}_i \) for each \( i \) is \( N(\mu_i, \Sigma_i(\alpha^R)) \) and hence is degenerate at \( \mu_i \) as \( z \to 0 \). Therefore
\[
\lim_{z \to 0} P_i(x_i | s_i) = \begin{cases} 
1 & x_i \geq \mu_i \\
0 & \text{otherwise}
\end{cases}
\]

and it follows that
\[
\lim_{z \to 0} P(x | \mu) = \begin{cases} 
1 & x \geq \mu \\
0 & \text{otherwise}
\end{cases},
\]

and hence the distribution of \( \tilde{x} \) is degenerate at \( \mu \).

**Theorem (4.2.14).** Let \( d \) be the decision rule given by the Heteroscedastic Method and suppose \( \exists N_1 \) such that \( R_{N_1}(\mu, d) \leq A(\mu) \) \( \forall \mu \in \Theta_1^* \) and \( A(\mu, d(\mu)) \leq A(\mu) \), then for each \( \mu \in \Theta_1^* \), \( \exists z_\mu \) such that \( R_{z_\mu}(\mu, d(\mu)) \leq A(\mu) \).

**Proof.** We need only show \( \lim_{z \to 0} R_z(\mu, d) = \ell(\mu, d(\mu)) \). \( R_z(\mu, d) = \int \ell(\mu, d(x)) dF(x | \mu) \) but from Corollary (4.2.13), \( F(x | \mu) \) is degenerate at \( \mu \) hence \( \lim_{z \to 0} R_z(\mu, d) = \ell(\mu, d(\mu)) < A(\mu) \). Thus if in the two-stage procedure we use the same decision rule \( d \) which meets our goal in the single-sample problem with \( \Phi \) known and equal for each population, at least pointwise we can select \( z \) so that the two-stage procedure also meets the requirement independently of the unknown \( \Phi_1, \ldots, \Phi_k \).
4.3. Examples of Some Statistical Decision Problems

Example (4.3.1). (Stein's original two-stage solution to Student's hypothesis). Let $k = 1$, $p = 1$ and $\pi_1 \sim N(\mu, \sigma^2)$. To test the hypothesis $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, we have action space $A = \{d_0, d_1\}$ where $d_0 = \text{accept}$ and $d_1 = \text{reject}$. $\Theta = \{ (\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0 \}$ and letting $w = \{ \mu : \mu = \mu_0 \}$ we define our loss function as

$$l(\mu, d(x)) = \begin{cases} 0, & d(x) = d_0 \text{ and } \mu \in w \\ 0, & d(x) = d_1 \text{ and } \mu \notin w \\ 1, & d(x) = d_0 \text{ and } \mu \notin w \\ 1, & d(x) = d_1 \text{ and } \mu \in w \end{cases}$$

Now to decide what decision rule $d$ to use in this case we apply (4.2.12). That rule will be a function of the sample mean $\bar{x}$ and $\frac{\sigma^2}{N}$ and we will use the same rule replacing $\bar{x}$ with the generalized sample mean $\tilde{x}$ and $\frac{\sigma^2}{N}$ with $z$ (note we take $(\alpha_{rs}) = (1)$ in this case). The rule we would use with $\sigma^2$ known would be

$$d(\bar{x}, \sigma^2/N) = \begin{cases} d_0 & \text{if } \left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{N}} \right| \leq c \\ d_1 & \text{if } \left| \frac{\bar{x} - \mu_0}{\sigma/\sqrt{N}} \right| > c \end{cases}$$

where $c$ is chosen to achieve a level $\alpha$ test. The rule we choose for the decision problem $(\Theta, A, \ell)$, $\bar{x}$ is therefore
where $c'$ is chosen to achieve a level $\alpha$ test. Thus we reject if and only if \( \left| \frac{\bar{x} - \mu_0}{\sqrt{z}} \right| > c' \) which is precisely the test proposed by Stein (1945) for this hypothesis.

**Example (4.3.2).** (The hypothesis testing problem proposed by Chapman.)

For this case $p = 1$, $k = 2$ (and again we take $(\alpha_{rs}) = (1)$). We want to test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ where $\pi_1 \sim N(\mu_1, \sigma_1^2)$ and $\pi_2 \sim N(\mu_2, \sigma_2^2)$. The loss function and action space are defined as in example 4.3.1 with $w = \{ (\mu_1, \mu_2) : \mu_1 = \mu_2 \}$. If $\sigma_1^2$ and $\sigma_2^2$ were known and equal the decision rule we would use based on single-stage samples of size $N$ would be

\[
d(\bar{x}, z) = \begin{cases} 
  d_0 & \text{if } \frac{\bar{x} - \mu_0}{\sqrt{z}} \leq c' \\
  d_1 & \text{if } \frac{\bar{x} - \mu_0}{\sqrt{z}} > c'. 
\end{cases}
\]

where $c'$ is chosen to achieve a level $\alpha$ test. Thus we reject if and only if \( \left| \frac{\bar{x} - \mu_0}{\sqrt{z}} \right| > c' \) which is precisely the test proposed by Stein (1945) for this hypothesis.

\[
d(\bar{x}, z) = \begin{cases} 
  d_0 & \text{if } \frac{\bar{x} - \mu_0}{\sqrt{2\sigma^2/N}} \leq c \\
  d_1 & \text{if } \frac{\bar{x} - \mu_0}{\sqrt{2\sigma^2/N}} > c
\end{cases}
\]

where $c$ is a fixed constant set to achieve a level $\alpha$ test. The rule we should use then would be
which is equivalent to Chapman's (1950) statistic.

**Example (4.3.3).** (The one-way layout with unequal variances.)

In section 2.1 we showed that our decision rule based on $F$ was derived from the considerations of section 4.2. Applying 4.2.12 if all the variances are known and equal we would take equal samples of size $N$ from each population and base our test on the statistic

$$F = \frac{k \sum_{i=1}^{k} \frac{N(x_i - \bar{x})^2}{\sigma^2}}{\sigma^2/n}$$

where $\sigma^2$ is the common known variance and $\bar{x} = 1/k \sum_{i=1}^{k} x_i$. Replacing $\bar{x}_i$ by $\hat{x}_i$ and $\sigma^2/N$ by $z$ we obtain the statistic $\hat{F}$.

**Example (4.3.4).** (Chatterjee's 1959 multivariate extension of Stein's work.) This is a special case of our decision problem with $k = 1$ and $p > 1$. $\pi \sim N_p(\mu, \Sigma)$ and we want to test $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$. The action space $A$ and loss function are the same as above but $\Theta = \{(\mu, \Sigma) : \mu \in \mathbb{R}^p, \Sigma$ is a positive-definite symmetric matrix$\}$ and $W = \{\mu : \mu = 0\}$. If $\Sigma$ were known we would use the likelihood ratio test statistic.
\[ U = N \bar{X}' \left( \frac{1}{N} \bar{X} \right)^{-1} \bar{X} = N \sum_{i=1}^{p} \sum_{j=1}^{p} \sigma_{ij} \bar{x}_i \bar{x}_j \]

where \( \bar{X} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p)' \) is the sample mean vector based on \( N \) observations and the decision rule would be

\[
d(\bar{X}, \frac{1}{N} \mathbb{I}) = \begin{cases} 
0 & \text{if } N \bar{X}' \left( \frac{1}{N} \bar{X} \right)^{-1} \bar{X} \leq c \\
1 & \text{if } N \bar{X}' \left( \frac{1}{N} \bar{X} \right)^{-1} \bar{X} > c
\end{cases}
\]

Following (4.2.12) for the case when \( \bar{X} \) is unknown we would use the decision rule

\[
d(\tilde{X}, z(\alpha^{rs})) = \begin{cases} 
0 & \text{if } \tilde{x}' (z(\alpha^{rs}))^{-1} \tilde{x} \leq c \\
1 & \text{if } \tilde{x}' (z(\alpha^{rs}))^{-1} \tilde{x} > c
\end{cases}
\]

That is, we would base the test on

\[ U = \tilde{x}' (z(\alpha^{rs}))^{-1} \tilde{x} = \frac{1}{z} \sum_{i=1}^{p} \sum_{j=1}^{p} \alpha_{ij} \tilde{x}_i \tilde{x}_j \]

since \( \tilde{x} = (x_1, x_2, \ldots, x_p)' \), which is exactly the statistic proposed by Chatterjee.

**Example (4.3.5).** (Multiple-comparison procedures.) As an example, consider the problem of simultaneously estimating all pairwise differences of \( k \) population means. In this case \( p = 1, k \geq 2 \),
\[ \Theta = \Theta_1 \times \Theta_2, \; \Theta_1 = \{ \theta : \theta \in \mathbb{R}^{k(k-1)/2} \}, \; \Theta_2 = \{ (\sigma_1^2, \ldots, \sigma_k^2) : \sigma_1^2 > 0 \}. \]

The loss function \( \ell \) may be defined as

\[
\ell(\theta, a) = \begin{cases} 
1, & \text{if } \theta \not\in d(\bar{\xi}) \\
0, & \text{if } \theta \in d(\bar{\xi}).
\end{cases}
\]

In the case when the variances are known and equal we would use the decision rule

\[
d(\bar{\xi}_1, \bar{\xi}_2, \ldots, \bar{\xi}_k, \sigma^2 / N) = \left\{ (\bar{\xi}_i - \bar{\xi}_j) \pm \frac{\sigma}{\sqrt{N}} t_{k,n}^\alpha, \; i, j = 1, 2, \ldots, k \right\}
\]

where \( t_{k,n}^\alpha \) is the upper \( \alpha \)th percent point of the range of \( k \) independent \( N(0,1) \) variates. Thus for the two-stage sampling scheme with unknown variances we should use

\[
d(\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_k, z) = \left\{ (\tilde{\xi}_i - \tilde{\xi}_j) \pm \sqrt{z} q_{k,n_0-1}^\alpha, \; i, j = 1, 2, \ldots, k \right\}
\]

where \( q_{k,n_0-1}^\alpha \) is the upper \( \alpha \) percent point of the range of \( k \) independent \( t_{n_0-1} \) variates.

Example (4.3.6). (The problem of selecting the best of \( k \) normal populations when the variances are unknown.) For this problem we have \( p = 1 \) and \( k \geq 2 \). \( \Theta = \Theta_1 \times \Theta_2, \; \Theta_1 = \{ \mu : \mu \in \mathbb{R}^k \}, \; \Theta_2 = \{ (\sigma_1^2, \ldots, \sigma_k^2) : \sigma_1^2 > 0 \} \).
and \( A = \{a_1, \ldots, a_k\} \) where \( a_i \) is the action of selecting population \( \pi_i \) as best. We define the loss function in this case as

\[
\ell((\mu_1, \mu_2, \ldots, \mu_k), d(\bar{X})) = \begin{cases} 
0, & \text{if } d(\bar{X}) = a_i \text{ and } \mu_i = \mu[k] \\
1, & \text{if } d(\bar{X}) = a_i \text{ and } \mu_i \neq \mu[k]
\end{cases}
\]

where \( \mu[k] = \max(\mu_1, \mu_2, \ldots, \mu_k) \). The decision rule one would use when the variances were known would be defined by (see Bechhofer (1954))

\[
d(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k, \sigma^2/N) = d_1 \quad \text{if } \bar{X}_i = \max(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k).
\]

Therefore the decision rule to use when the variances are unknown is (by 4.2.12)

\[
d(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k, z) = d_1 \quad \text{if } \bar{X}_i = \max(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_k)
\]

which is precisely the rule of Dudewicz and Dalal (1975).

The function of \( z \) playing the role of \( \sigma^2/N \) is not as clear in this case. However, in the ranking and selection problem it is desired to detect a minimum distance \( \delta^* \) between \( \mu[k-1] \) and \( \mu[k] \) where \( \mu[1] \leq \mu[2] \leq \cdots \leq \mu[k] \) denote the ordered means and \( \sigma^2/N \) (i.e. \( N \)) is chosen so that

\[
\int_{-\infty}^{\infty} \phi^{k-1}(x + \frac{\delta^* \sqrt{N}}{\sigma}) \varphi(x) \, dx \geq p^*
\]
where $P^*$ is a fixed constant $1/k < P^* < 1$, $\Phi$ is the $N(0,1)$ CDF and $\varphi$ the $N(0,1)$ density.

Similarly the two-stage sampling scheme for ranking and selection is defined by

$$N_1 = \max \{ n_0 + 1, [s_1^0/z + 1] \}$$

and $z$ is chosen so that

$$\int_{-\infty}^{\infty} F_{n_0}^{-1}(X - \delta^*/\sqrt{z}) f_{n_0}(x) \, dx \geq P^*$$

where $F_{n_0}$ and $f_{n_0}$ are the distribution and density functions respectively for a $t_{n_0-1}$ variate.

Thus as before the two-stage decision rule and choice of $z$ follow from the general decision theory guidelines. Note that rule is exactly that of Dudewicz and Dalal (1975); Dudewicz and Dalal chose $h$ to satisfy

$$\int_{-\infty}^{\infty} F_{n_0}^{-1}(x+h) f_{n_0}(x) \, dx \geq P^*$$

and define

$$N_1 = \max \{ n_0 + 1, [s_1^0/\delta^* + 1] \} ,$$

which is equivalent to the above with $z = (\delta^*/h)^2$. 
Nonexistence of Single-Sample Procedures

Dantzig (1940), Stein (1945), Dudewicz (1971), and Dudewicz and Dalal (1975) have proven that there do not exist single-sample procedures which satisfy the requirements of their respective problems. In this section we consider the general decision problem and show that, under certain conditions on the risk function, no single-sample procedure exists which can meet our goals. The assumptions on the risk function include all of the above cases as will be shown.

Let \((\Theta, A, s), X\) be the decision problem given by (4.1.1) except let \(X = (\bar{X}_1, \ldots, \bar{X}_k)\) where \(\{\bar{X}_i\}_{i=1}^k\) are independent \(\bar{X}_i\) being the mean vector of a sample based on \(N_i\) observations from \(\pi_i \sim N(\mu_i, \sigma_i)\), \(i = 1, \ldots, k\), since \(X\) is a sufficient statistic for the single sample problem. Thus we are considering a single-sample decision problem. Let \(R(\theta, s)\) denote the risk function associated with the single-stage decision rule \(s\). Our

Goal: to find a rule \(s\) such that uniformly in \(\theta_1, \ldots, \theta_k\),

\[R(\theta, s) \leq A(\theta) \quad \forall \theta \in \Theta^* \times \Theta_2, \Theta^* \subseteq \Theta_1\]  \hfill (4.4.1)

where \(A(\theta)\) is a fixed constant for each \(\theta\). Letting \(C\) denote the class of all single-sample decision rules, we have the following result.

Theorem (4.4.2). If \(\inf_{s \in C} R(\theta, s) = g(\theta, \theta_1, \ldots, \theta_k)\) where

\[\sup_{\theta \in \Theta^*} g(\theta, \theta_1, \ldots, \theta_k) = B(\theta)\]  \hfill (4.4.2)

and if there exists at least one \(\theta \in \Theta_1\) such that \(B(\theta^*) > A(\theta^*)\) then no single-stage procedure exists which satisfies goal (4.4.1).
Proof. Suppose $s^* \in C$ satisfies (4.4.1). Then

$$R(\theta^*, s^*) \geq \inf_{s \in C} R(\theta^*, s) = g(\theta^*, \xi_1, \ldots, \xi_k)$$

so

$$\sup_{\xi_1, \ldots, \xi_k} R(\theta^*, s^*) \geq \sup_{\xi_1, \ldots, \xi_k} g(\theta^*, \xi_1, \ldots, \xi_k) = B(\theta^*) > A(\theta^*)$$

which is a contradiction, hence no such rule exists.

We now show that the problems considered by Dantzig, Stein, Dudewicz, and Dudewicz and Dalal are special cases of Theorem (4.4.2).

Example (4.4.3). Since Dantzig's problem is a special case of Stein's we consider only Stein's. In this case $N_1 = N$, $k \geq 1$, $p = 1$ and $\pi_i \sim N(u_i, \sigma^2)$, $i = 1, 2, \ldots, q$, $\pi_I \sim N(0, \sigma^2)$, $I = q+1, \ldots, k$ (canonical form). $\Theta^* = \{\theta_1 = (0, 0, \ldots, 0), \xi_2 = (u_{10}, u_{20}, \ldots, u_{q0}, 0, \ldots, 0)\}$. $A$ and $\ell$ are defined as in example 4.2.1 where $w = \{(u_{10}, u_{20}, \ldots, u_{q0}, 0, \ldots, 0)\}$. Now for this problem $R(\theta_1, s) = \alpha$ the level of the test and $R(\theta_2, s) = \beta$, the probability of a type II error. Our goal then is to have $R(\theta_2, s) \leq \beta^*$ where $\beta^*$ is some constant $0 < \beta^* < 1 - \alpha$ and we take $\theta^* = w$. The class $C$ is all single-stage decision rules and $s^*$ is the decision rule given by the Neyman-Pearson lemma. (i.e. the most powerful level $\alpha$ test of $H_0 : u_1 = 0, \ldots, u_q = 0$ versus $H_1 : u_1 = u_{10}, \ldots, u_q = u_{q0}$). It then follows that
\[ R(\theta_2, s^*) = 1 - \frac{\sum_{i=1}^{q} \mu_i^2}{\left(\frac{\sigma}{\sqrt{n}}\right) \sqrt{\sum_{i=1}^{q} \mu_i^2}} \cdot c - \sqrt{\frac{\sum_{i=1}^{q} \mu_i^2}{\sigma/\sqrt{n}}} = g(\sigma^2). \]

It is clear that \( \sup_{\sigma^2} g(\sigma^2) = 1 - \alpha > \beta^* \) hence by Theorem (4.4.2) no single stage procedure exists which will satisfy our goal uniformly in \( \sigma^2 \).

**Example (4.4.4).** Consider the example of selecting the best of \( k \) normal populations when the variance are equal and unknown. In this case the decision problem is \((\Theta, A, \ell), X\) with \((\Theta, A, \ell)\) defined as in Example 4.2.6 and \( X = (\tilde{X}_1, \ldots, \tilde{X}_k) \) where \( \tilde{X}_i \) are independent sample means based on samples of size \( N \) from \( \pi_i \sim \text{N}(\mu_i, \sigma^2) \). The risk function for this problem is \( R(\ell, s) = 1 - P[CS] \) and our goal is to find a rule \( s \) such that \( R(\theta, s) \leq 1 - P^* \) where \( 0 < P^* < 1/k \) or \( 0 < 1 - P^* < 1 - 1/k \) for \( \theta \in \Theta^* \) where

\[ \Theta^* = \{ (\mu_1, \ldots, \mu_k) : \mu_{[k]} - \mu_{[k-1]} \geq \delta^*, \delta^* > 0 \} \]  

Hall (1959) showed that the optimal decision rule \( s^* \) is this case was the one developed by Bechhofer (1954). In that case

\[ R(\theta, s^*) = 1 - \int_{-\infty}^{\infty} \phi \left( y + \frac{\mu_{[k]} - \mu_{[l]}}{\sigma/\sqrt{N}} \right) \cdots \phi \left( y + \frac{\mu_{[k]} - \mu_{[k-1]}}{\sigma/\sqrt{N}} \right) \varphi(y) \, dy \]

\[ = g(\sigma^2) \]
and \( \sup_{\sigma^2} g(\sigma^2) = 1 - 1/k > 1-P^* \), hence no single-stage procedure can guarantee the probability requirement uniformly in \( \sigma^2 \). (Note proofs similar to these apply to other goals as in Dudewicz and Dalal (1975). Also the normality assumption in Theorem 4.4.2 is not really required and thus the theorem is more general than stated.)
4.5. Limiting Distribution of $\mathbf{Y}$

We have shown that for the statistical decision problem $(\Theta, \mathcal{A}, \ell)$, the associated risk function $R(\mu, d) = \int \ell(\mu, d(\mathbf{y})) \, dF_{\mathbf{y}/\mu}$ is independent of the unknown covariance matrices and controllable in the sense of Theorem (4.2.14). However, to use (4.2.12) in practice we must be able to evaluate $R(\mu, d)$. For Example, 4.2.1, we know that the risk function gives the probabilities of Type I and Type II errors and in that case $\mathbf{Y}$ reduces to a generalized sample mean from a univariate population and hence follows a $t_{n_0-1}$ distribution, thus $R(\mu, d)$ was easily evaluated. Similarly Dudewicz and Dalal were able to numerically evaluate $R(\mu, d)$ for the selection problem. However, in other cases (such as Example 4.2.3) the distribution of $\mathbf{Y}$ has not been calculated so $R(\mu, d)$ has not yet been tabled exactly and an approximation is needed until such tabling is done. We consider approximating $R(\mu, d)$ by its limiting value as $n_0 \to \infty$.

**Theorem (4.5.1).** If the loss function $\ell(\mu, d)$ is a bounded continuous function of $d$ then

$$
\lim_{n_0 \to \infty} R(\mu, d) = \int \ell(\mu, d(y)) \, dG(y)
$$

where $G(y)$ is the distribution function of $Y = (Y_1, \ldots, Y_k)'$, where $\{Y_i\}_{i=1}^k$ are independent $Y_i \sim N_p(\mu_i, \Sigma \alpha^p)$.
Proof. Using an argument similar to that for Theorem (2.6.41) it can be shown that as \( n_0 \to \infty \) the limiting distribution of \( \bar{X}_i \) is \( N_p(\mu_i, z_p(\alpha^RS)) \). Therefore the limiting distribution of \( \bar{X} = (\bar{X}_1, \ldots, \bar{X}_k) \) is the same as that of \( \bar{Y} \). If \( G_n(y) \) is the distribution function for \( \bar{X} \) then

\[
R(\mu, d) = \int \ell(\mu, d(y)) dG_{n_0}(y)
\]

and it follows from the Helly-Bray Theorem (see page 84 of Tucker (1967)) that

\[
\lim_{n_0 \to \infty} R(\mu, d) = \int \ell(\mu, d(y)) dG(y) .
\]
5.1. Introduction

In Chapter 4 we developed a general theory which allows us to handle any statistical decision problem concerning the mean vectors of \( k \) \( p \)-dimensional multivariate normal distributions with unknown and unequal covariance matrices. In this chapter we apply this to the multivariate analysis of variance (MANOVA) in the one-way layout (and some related multiple-comparison procedures), areas sorely in need of such an analysis.
5.2. The One-Way MANOVA Layout

Let $\pi_1, \pi_2, \ldots, \pi_k$ be $k$ independent populations such that
$\pi_i \sim N_p(\mu_i, \Sigma)$ where $\mu_i$ and $\Sigma$ are unknown. To test the null hypothesis

$$H_0: \mu_1 = \cdots = \mu_k$$

(5.2.1)

we use the two-stage sampling scheme developed in Chapter 4 and choose our test statistic according to the Heteroscedastic Method (4.2.12). In order to decide on our test statistic we must first consider the statistic we would use if we were sampling from $\pi_i \sim N_p(\mu_i, \Sigma)$ (i = 1, 2, ..., k) and $\Sigma$ were known. In that case we would take samples of size $N$ from each population and base the test of (5.2.1) on the likelihood ratio test statistic $\lambda$.

**Lemma (5.2.2).** Let $\pi_1, \pi_2, \ldots, \pi_k$ be $k \geq 2$ p-dimensional multivariate normal populations such that $\pi_i \sim N_p(\mu_i, \Sigma)$ where $\Sigma$ is known. The likelihood ratio test of (5.2.1) is based on the statistic

$$\lambda = N \sum_{i=1}^{k} (\bar{X}_i - \bar{X})' \Sigma^{-1} (\bar{X}_i - \bar{X})$$

where

$$\bar{X} = \frac{1}{k} \sum_{i=1}^{k} \bar{X}_i.$$

**Proof.** Let $X_{i1}, X_{i2}, \ldots, X_{iN}$ be the random sample from $\pi_i$. Then the likelihood function in general is
\[
L(\mu_1, \ldots, \mu_k) = \prod_{k=1}^{N} \prod_{j=1}^{N} \frac{(2\pi)^{-p/2} |\Phi|^{-1/2}}{e} - \frac{1}{2} \sum_{i,j} (X_{ij} - \mu_i)' \Phi^{-1}(X_{ij} - \mu_i) \\
= (2\pi)^{-kN p/2} |\Phi|^{-kN/2} e - \frac{1}{2} \sum_{i,j} \sum_{i,j} (X_{ij} - \mu_i)' \Phi^{-1}(X_{ij} - \mu_i)
\]

and

\[
\log L(\mu_1, \ldots, \mu_k) = - \frac{kN p}{2} \log 2\pi - \frac{kN}{2} \log |\Phi| - \frac{1}{2} \sum_{i,j} \sum_{i,j} (X_{ij} - \mu_i)' \Phi^{-1}(X_{ij} - \mu_i).
\]

\[
\log L \text{ is maximized by minimizing}
\[
\sum_{i=1}^{k} \sum_{j=1}^{N} (X_{ij} - \mu_i)' \Phi^{-1}(X_{ij} - \mu_i)
\]

Now,

\[
\sum_{i=1}^{k} \sum_{j=1}^{N} (X_{ij} - \mu_i)' \Phi^{-1}(X_{ij} - \mu_i)
\]

\[
= \text{tr} \sum_{i=1}^{k} \sum_{j=1}^{N} (X_{ij} - \mu_i)' \Phi^{-1}(X_{ij} - \mu_i)
\]

\[
= \text{tr} \sum_{i=1}^{k} \sum_{j=1}^{N} [/ \Phi^{-1}(X_{ij} - \mu_i)(X_{ij} - \mu_i)' ]
\]

\[
= \text{tr} \Phi^{-1} [ \sum_{i=1}^{k} \sum_{j=1}^{N} (X_{ij} - \bar{X}_i + \bar{X}_1 - \mu_i)(X_{ij} - \bar{X}_i + \bar{X}_1 - \mu_i)' ]
\]

\[
= \Phi^{-1} [ \sum_{i=1}^{k} \sum_{j=1}^{N} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)' + \sum_{i=1}^{k} \sum_{j=1}^{N} (\bar{X}_1 - \mu_1)(X_{ij} - \bar{X}_i)' + \sum_{i=1}^{k} \sum_{j=1}^{N} (\bar{X}_1 - \mu_1)(\bar{X}_1 - \mu_1)' ]
\]
\[
\begin{align*}
&= \text{tr} \, \frac{1}{2} \left[ \sum_{i} \sum_{j} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)' + \frac{N}{k} \sum_{i} (\bar{x}_i - \bar{\mu}_i)(\bar{x}_i - \bar{\mu}_i)' \right] \\
&= \frac{k}{N} \sum_{i} \sum_{j} (x_{ij} - \bar{x}_i)' \frac{1}{2} (x_{ij} - \bar{x}_i) + \frac{N}{k} \sum_{i} (\bar{x}_i - \bar{\mu}_i)' \frac{1}{2} (\bar{x}_i - \bar{\mu}_i)
\end{align*}
\]

Since \( \frac{1}{2} \) is positive-definite, we minimize \( \sum_{j} \sum_{j} (x_{ij} - \bar{\mu}_i)(x_{ij} - \bar{\mu}_i)' \) by setting \( \bar{\mu}_i = \bar{x}_i \). Under 5.2.1 we want to maximize

\[
L(y) = \prod_{i=1}^{k} \prod_{j=1}^{N} (2\pi)^{-p/2} |y|^{-1/2} e^{-\frac{1}{2} (x_{ij} - y)' \frac{1}{2} (x_{ij} - y)}
\]

It follows in the same manner as above that \( L(y) \) is maximized when \( \mu = \bar{x} = 1/k \sum_{i=1}^{k} \bar{x}_i \). The likelihood-ratio test is thus based on

\[
\lambda = \max_{\mu} \frac{L(\mu)}{\max_{\mu_1, \cdots, \mu_k} L(\mu_1, \cdots, \mu_k)}
\]

\[
\begin{align*}
\lambda &= e^{-\frac{1}{2} \sum_{i} \sum_{j} (x_{ij} - \bar{x}_i)' \frac{1}{2} (x_{ij} - \bar{x}_i) - \frac{1}{2} \sum_{i} \sum_{j} (x_{ij} - \bar{x}_1)' \frac{1}{2} (x_{ij} - \bar{x}_1)} \\
&= e \left[ \frac{1}{2} \sum_{i} \sum_{j} (x_{ij} - \bar{x}_i)' \frac{1}{2} (x_{ij} - \bar{x}_i) - (x_{ij} - \bar{x}_1)' \frac{1}{2} (x_{ij} - \bar{x}_1) \right]
\end{align*}
\]

Hence

\[
\log \lambda = \frac{1}{2} \sum_{i} \sum_{j} \left[ (x_{ij} - \bar{x}_i)' \frac{1}{2} (x_{ij} - \bar{x}_i) - (x_{ij} - \bar{x}_1)' \frac{1}{2} (x_{ij} - \bar{x}_1) \right]
\]

Now
\[
2kO
k \quad k = t \quad \sigma \quad \sigma 
[272x684] \quad [292x684] \\
= \quad t \quad \sigma \quad \sigma 
[322x684] \quad \sigma \quad \sigma 
\]

Since \( \Psi^{-1} \) is positive-definite, we minimize \( \frac{1}{k} \sum \sum (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)' \) by setting \( \mu_i = \bar{x}_i \). Under 5.2.1 we want to maximize

\[
L(\mu) = \prod_{i=1}^{k} \prod_{j=1}^{N} (2\pi)^{-p/2} |\Psi|^{-1/2} e^{-\frac{1}{2}(x_{ij} - \mu_i)' \Psi^{-1} (x_{ij} - \mu_i)}
\]

It follows in the same manner as above that \( L(\mu) \) is maximized when \( \mu = \bar{X} = 1/k \sum_{i=1}^{k} \bar{x}_i \). The likelihood-ratio test is thus based on

\[
\max_{\hat{\mu}} L(\hat{\mu})
\]

\[
\lambda = \frac{\max_{\mu_1} L(\mu_1)}{\max_{\mu_1} L(\mu_1, \ldots, \mu_k)}
\]

\[
-\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{N} (x_{ij} - \bar{x}_i)' \Psi^{-1} (x_{ij} - \bar{x}_i)
\]

\[
= e^{-\frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{N} (x_{ij} - \bar{x}_i)' \Psi^{-1} (x_{ij} - \bar{x}_i)}
\]

\[
+ \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{N} [(x_{ij} - \bar{x}_i)' \Psi^{-1} (x_{ij} - \bar{x}_i) - (x_{ij} - \bar{x}_i)' \Psi^{-1} (x_{ij} - \bar{x}_i)]
\]

Hence

\[
\log \lambda = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{N} [(x_{ij} - \bar{x}_i)' \Psi^{-1} (x_{ij} - \bar{x}_i) - (x_{ij} - \bar{x}_i)' \Psi^{-1} (x_{ij} - \bar{x}_i)]
\]

Now
\[\sum_{i<j} (x_{ij} - \bar{x}_1)^t \frac{1}{n} (x_{ij} - \bar{x}_1)\]

\[= \sum_{i<j} (x_{ij} - \bar{x} + \bar{x} - \bar{x}_1)^t \frac{1}{n} (x_{ij} - \bar{x} + \bar{x} - \bar{x}_1)\]

\[= \sum_{i<j} (x_{ij} - \bar{x})^t \frac{1}{n} (x_{ij} - \bar{x}) + \sum_{i<j} (x_{ij} - \bar{x})^t \frac{1}{n} (\bar{x} - \bar{x}_1)\]

\[+ \sum_{i<j} (\bar{x} - \bar{x}_1)^t \frac{1}{n} (x_{ij} - \bar{x}) + \sum_{i<j} (\bar{x} - \bar{x}_1)^t \frac{1}{n} (\bar{x} - \bar{x}_1)\]

Further

\[\sum_{i<j} (x_{ij} - \bar{x})^t \frac{1}{n} (x_{ij} - \bar{x}) = \sum_{i<j} \frac{1}{n} (x_{ij} - \bar{x})^t (x_{ij} - \bar{x})\]

\[= \frac{k}{n} \sum_{i} (x_i - \bar{x})^t \frac{1}{n} (x_i - \bar{x}) = \frac{k}{n} \sum_{i} (\bar{x}_i - \bar{x})^t \frac{1}{n} (\bar{x}_i - \bar{x})\]

Similarly,

\[\sum_{i<j} (\bar{x} - \bar{x}_1)^t \frac{1}{n} (x_{ij} - \bar{x}) = \frac{k}{n} \sum_{i} (\bar{x}_i - \bar{x})^t \frac{1}{n} (\bar{x}_i - \bar{x})\]

Hence

\[\log \lambda = -\frac{n}{2} \sum_{i} \frac{k}{n} (\bar{x}_i - \bar{x})^t \frac{1}{n} (\bar{x}_i - \bar{x})\]

and hence we may equivalently base our test on the statistic

\[N \sum_{i=1}^{k} (\bar{x}_i - \bar{x})^t \frac{1}{n} (\bar{x}_i - \bar{x})\]. Following the guidelines of Chapter 4, we shall use the two-stage sampling scheme and base our test on the statistic
where \( \bar{X} = \frac{1}{k} \sum_{i=1}^{k} \bar{x}_i \), \( z > 0 \) and \( (\alpha_{rs}) \) are the prespecified positive constant and positive definite matrix in the sampling scheme, and \( \bar{x}_i \) are the generalized mean vectors.

**Lemma (5.2.4).** The distribution of \( F \) is independent of \( \mu_1 \) under both null and alternative hypotheses.

**Proof.** Follows directly from the general theory of Chapter 4.

The test proceeds by rejecting \( H_0 : \mu_1 = \cdots = \mu_k \) if and only if \( \bar{F} > C_{p,k,\alpha} \) where \( C_{p,k,\alpha} \) is the upper \( \alpha \)th percent point of the null distribution of \( F \).
5.3. The Limiting Distribution of \( F \)

The exact distribution of \( F \) is not tabulated and some approximation is needed. As before we consider the limiting distribution of \( F \) (as \( n_0 \to \infty \)) as an approximation to the true distribution. In order to derive this limiting distribution we need the following lemmas.

(Recall that \( A \otimes B \) denotes the direct product of definition (1.2.4).)

**Lemma (5.3.1).** If \( A \) is a \( q \times q \) matrix of rank \( r \) and \( B \) is a \( p \times p \) matrix of rank \( s \) then the rank of \( A \otimes B \) is \( rs \).

**Proof.** From Anderson (1958), pg. 347 if \( \lambda_i \) is the \( i \)th characteristic root of \( A \) and \( r_\alpha \) is the \( \alpha \)th characteristic root of \( B \) then \( \lambda_i r_\alpha \) is the \( i, \alpha \)th characteristic root of \( A \otimes B \). Now there are \( r \) nonzero \( \lambda_i \) and \( s \) nonzero \( r_\alpha \) and so there are \( rs \) nonzero characteristic roots of \( A \otimes B \) hence the rank of \( A \otimes B \) is \( rs \).

**Lemma (5.3.2).** If \( A \) is a \( q \times q \) idempotent matrix of rank \( r \) and \( I \) is the \( p \times p \) identity matrix then \( A \otimes I \) is an idempotent matrix of rank \( p \cdot r \).

**Proof.** A idempotent implies \( a_{ij} = \sum_{k=1}^{q} a_{ik} a_{kj}, \) \( i,j = 1, \ldots, q \)

\[
A \otimes I = \begin{bmatrix}
a_{11} I & a_{12} I & \cdots & a_{1q} I \\
a_{21} I & a_{22} I & \cdots & a_{2q} I \\
\vdots & \vdots & \ddots & \vdots \\
a_{q1} I & a_{q2} I & \cdots & a_{qq} I
\end{bmatrix}
\]
so

\[(A \otimes I)(A \otimes I) = \left[ \sum_{j=1}^{q} a_{ij} b_{ij} I \right] = \left[ a_{ij} I \right].\]

The rank of \( A \otimes I \) follows from Lemma (5.3.1).

**Lemma (5.3.3).** Let \( y_i = (y_{i1}, \ldots, y_{ip})' \) then

\[\sum_{i=1}^{k} (y_i - \bar{y})'(y_i - \bar{y}) = \left[ \begin{array}{cccc} y_{11}' & \cdots & y_{p1}' \\ \vdots & \ddots & \vdots \\ y_{1k}' & \cdots & y_{pk}' \end{array} \right] A \otimes I \left[ \begin{array}{c} y_{11} \\ \vdots \\ y_{1k} \end{array} \right]
\]

where \( A = I_{k \times k} - \frac{1}{k} I \), \( I_{k \times k} \) is the \( k \times k \) identity, \( I \) is the \( p \times p \) identity, and \( \bar{y} = \frac{1}{k} \sum_{i=1}^{k} y_i \).

**Proof.**

\[
\left[ \begin{array}{cccc} y_{11}' & \cdots & y_{p1}' \\ \vdots & \ddots & \vdots \\ y_{1k}' & \cdots & y_{pk}' \end{array} \right] \left[ \begin{array}{cccc} (1-1/k)I & -1/kI & -1/kI \\ -1/kI & (1-1/k)I & -1/kI \\ \vdots & \vdots & \vdots \\ -1/kI & -1/kI & (1-1/k)I \end{array} \right] \left[ \begin{array}{c} y_{1} \\ y_{2} \\ \vdots \\ y_{k} \end{array} \right]
\]

\[= \left[ \begin{array}{c} y_{1}'(1-1/k) - 1/k \sum_{j \neq 1} y_{j}' \ldots, y_{k}'(1-1/k) - 1/k \sum_{j \neq k} y_{j}' \end{array} \right] \]

\[= \frac{k}{i=1} \sum (y_i - \bar{y}) y_i = \frac{k}{i=1} \sum (y_i - \bar{y})'(y_i - \bar{y}).\]
Now consider the random variable

$$Q = \sum_{i=1}^{k} (\bar{Y}_i - \bar{Y})' V^{-1}(\bar{Y}_i - \bar{Y})$$

where

$$\bar{Y} = \frac{1}{k} \sum_{i=1}^{k} Y_i$$

and the $\bar{Y}_i$ are independent $Y_i \sim N_p(\mu_i, V)$.

Lemma (5.3.4). The distribution of $Q$ is noncentral chi-square with $p(k-1)$ degrees of freedom and noncentrality parameter

$$\sum_{i=1}^{k} (\mu_i - \bar{\mu})' V^{-1}(\mu_i - \bar{\mu}) .$$

Proof. Let $Y = [Y_1', Y_2', \ldots, Y_k']$ then $Y \sim N_{pk}(\mu, V)$ where $\bar{\mu} = [\mu_1', \ldots, \mu_k']$ and

$$V = \begin{bmatrix} V & 0 & \cdots & 0 \\ 0 & V \\ \vdots & \vdots \\ 0 & \cdots & V \end{bmatrix}$$

Now $V$ is positive definite so there exists a nonsingular matrix $C$ such that $C V C' = I$. Thus,
\[
Q = \sum_{i=1}^{k} (y_i - \bar{Y})'(V^{-1}(y_i - \bar{Y})
\]
\[
= \sum_{i=1}^{k} (y_i - \bar{Y})'c'(c')^{-1}V^{-1}c^{-1}c(y_i - \bar{Y})
\]
\[
= \sum_{i=1}^{k} (Cy_i - Cy\bar{y})'(Cy_i - Cy\bar{y})
\]
\[
= [(Cy_1)',\ldots,(Cy_k)] A \otimes I[(Cy_1)',\ldots,(Cy_k)']
\]

where \( A \) is the \( k \times k \) matrix defined in Lemma (5.3.3). Now let

\[
X = [(Cy_1)',\ldots,(Cy_k)']
\]

then

\[
X \sim N_{pk}(\bar{C}y, I) \quad \text{where} \quad \bar{C} = \begin{bmatrix}
C & 0 & \cdots & 0 \\
0 & C & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & C
\end{bmatrix}
\]

and \( I \) is the \( pk \times pk \) identity matrix. From Lemma (5.3.2) \( A \otimes I \) is an idempotent matrix of rank \( p(k-1) \) since \( A \) is idempotent and has rank \( k-1 \). \( Q = X'(A \otimes I)X \) and from Lemma (2.1.17) it follows that \( Q \) is distributed as a noncentral chi-square with \( p(k-1) \) degrees of freedom and noncentrality parameter \( (\bar{C}y)'(A \otimes I)(\bar{C}y) \)
\[
\begin{align*}
&= \sum_{i=1}^{k} (\mu_i - \bar{\mu})'(C \mu_i - \bar{C} \mu) \\
&= \sum_{i=1}^{k} (\mu_i - \bar{\mu})' C C (\mu_i - \bar{\mu}) \\
&= \sum_{i=1}^{k} (\mu_i - \bar{\mu})' V^{-1} (\mu_i - \bar{\mu}).
\end{align*}
\]

We may now derive the limiting distribution of \( \tilde{F}/p \).

**Corollary (5.3.5).** The limiting distribution of \( \tilde{F}/p \) is noncentral chi-square with \( p(k-1) \) degrees of freedom and noncentrality parameter

\[
\frac{1}{pz} \sum_{i=1}^{k} (\mu_i - \bar{\mu})' (\alpha_{rs})(\mu_i - \bar{\mu}).
\]

**Proof.** We have shown in Chapter 4 that the limiting distribution of \( \tilde{X}_1^2 \) as \( n_0 \to \infty \) is \( N_p (\mu_1, Zp(\alpha_{rs})) \). \( \tilde{F}/p \) is a continuous function of the \( \tilde{X}_1 \) and hence in the limit is distributed as \( Q \) in Lemma (5.3.4) with \( V = Zp(\alpha_{rs}) \).

Comparison of this two-stage test procedure with the "usual" MANOVA techniques will be dealt with in future research.
5.4. Multivariate Multiple Comparisons--One Population

As in the univariate case if we are testing hypotheses about the mean vector of a single multivariate normal population or about the mean vectors of several populations if we reject we want to investigate the cause of rejection. In this section we consider simultaneous inference techniques for one or several mean vectors which yield confidence intervals for linear combinations of mean vector components which are of fixed length. We begin by considering the case $k = 1$ and $p > 1$ which we have noted earlier was considered by Chatterjee (1959a) and Healy (1956).

Let $\pi \sim N_p(\mu, \Sigma)$ where $\mu$ and $\Sigma$ are unknown. A test of $H_0 : \mu = \mu_0$ may be constructed which is independent of $\Sigma$ based on the two-stage sampling scheme of Chapter 4 and the test statistic

$$U = (\tilde{X} - \mu_0)' \left[ 2\sigma_p^2 \right]^{-1} (\tilde{X} - \mu_0) \tag{5.4.1}$$

where the distribution of $U$ is given by (1.3.20). If we reject $H_0$ we would like to determine what components or combination of components of $\mu$ are causing the rejection. We consider a family of simultaneous confidence intervals for all linear combinations of the components of $\mu$ (note that Chatterjee found a confidence region for $\mu$ of fixed dimensions).

Theorem (5.4.2). The probability is $1-\alpha$ that all linear combinations $l'\mu$, $l \in \mathbb{R}^p$, are simultaneously covered by the family of intervals
where \( u_{p, \alpha} \) is the upper \( \alpha \)th percent point of the distribution of \( U \).

**Proof.** \( \alpha^{rs} \) is positive definite so there exists a nonsingular matrix \( P \) such that \( P[\alpha^{rs}]P' = I \). Let \( \ell = P' \lambda \) then

\[
|\ell' (\tilde{X} - \mu)| \leq \left( u_{p, \alpha} \right)^{1/2} [\ell' (\alpha^{rs}) \ell]^{1/\alpha} \quad \forall \ell \in \mathbb{R}^p
\]

if and only if

\[
|\lambda' P(\tilde{X} - \mu)| \leq \left( u_{p, \alpha} \right)^{1/2} [\lambda' P(\alpha^{rs})P' \lambda]^{1/2} \quad \forall \lambda \in \mathbb{R}^p
\]

By Lemma (2.2.8)

\[
|\lambda' (\tilde{P} \tilde{X} - \tilde{P} \mu)| \leq \left( u_{p, \alpha} \right)^{1/2} [\lambda' \lambda]^{1/2} \quad \forall \lambda \in \mathbb{R}^p
\]

if and only if

\[
(\tilde{P} \tilde{X} - \tilde{P} \mu)'(\tilde{P} \tilde{X} - \tilde{P} \mu) = (\tilde{P} \mu)' P' P (\tilde{P} \mu) = (\tilde{X} - \mu)(\alpha^{rs}) (\tilde{X} - \mu) \leq u_{p, \alpha}
\]

since \( P' P = [\alpha^{rs}] \). But,

\[
P[(\tilde{X} - \mu)' (\alpha^{rs})(\tilde{X} - \mu) \leq u_{p, \alpha}] = P[\tilde{X}' [z^{rs} \lambda]^{1/\lambda} \tilde{X} \leq u_{p, \alpha}] = 1 - \alpha.
\]
The upper $\alpha$th percent point $u_{p,\alpha}$ is unknown except for $p = 2$ (see Chatterjee (1959a)) and thus we need to approximate $u_{p,\alpha}$. We note that the limiting distribution of $U/p$ is $\chi^2(p)$ and thus the probability is approximately $1-\alpha$ that all linear combinations $z'\bar{y}$ are covered by the family

$$\{ z'\frac{F}{p,\alpha} + (z^{2}) \}^{1/2} [z'(\alpha_{rs}z)]^{1/2}$$

(5.4.3)
5.5. Multiple Comparison Procedures with MANOVA

In this section we consider multiple comparison procedures which may be used if the hypothesis 5.2.1 is rejected. The parameters of interest in such a case are the linear combinations

\[ \mu_{c',d'} = \sum_{h=1}^{k} d_h c' \xi_h \]  

(5.5.1)

where \( c' = (c_1, \ldots, c_p) \in \mathbb{R}^p \) and \( d' = (d_1, \ldots, d_k) \in \mathbb{R}^k \) and \( c',d' \) are arbitrary.

**Theorem (5.5.2).** The probability is 1-\( \alpha \) that all the linear combinations \( \mu_{c',d'} \) are fixed, are simultaneously covered by the family of intervals

\[ \left| \sum_{h=1}^{k} d_h c' \tilde{\xi}_h \right| \leq \xi_{k,p,\alpha} \left( \sum_{h=1}^{k} d_h^2 \right)^{1/2} \left( c' (p z^r c) \right)^{1/2} \]

where \( \xi_{k,p,\alpha} \) is the upper \( \alpha \)th percent point of the distribution of

\[ \sum_{h=1}^{k} \left[ \frac{c' (\tilde{\xi}_h - \mu_h)}{\left( c' (p z^r c) \right)^{1/2}} \right]^2 \]

**Proof.** Let \( y_h = \frac{c' (\tilde{\xi}_h - \mu_h)}{\left( c' (p z^r c) \right)^{1/2}} \), then from Lemma (2.2.8)

\[ \left| \sum_{h=1}^{k} d_h y_h \right| \leq \xi_{k,p,\alpha} \left( \sum_{h=1}^{k} d_h^2 \right)^{1/2} \]

if and only if

\[ \sum_{h=1}^{k} y_h^2 < \xi_{k,p,\alpha} \]
Since
\[
\sum_{h=1}^{k} d_{h} \bar{y}_{h} = \sum_{h=1}^{k} d_{h} \frac{c' \bar{x}_{h} - \bar{\mu}_{h}}{\sqrt{c'(pz^{rs})c}}
\]
the result follows.

As before the critical value \( \xi_{k,p,\alpha} \) is not tabulated and so we approximate it by the limiting distribution of
\[
\sum_{h=1}^{k} \frac{(c' \bar{x}_{h} - \bar{\mu}_{h})^2}{\sqrt{c'(pz^{rs})c}}
\]
as \( n_0 \to \infty \). We know \( \frac{c' \bar{x}_{h} - \bar{\mu}_{h}}{\sqrt{c'(pz^{rs})c}} \) hence \( c' \bar{x}_{h} - \bar{\mu}_{h} \) \( \to N(0, c'(pz^{rs})c) \) and so
\[
\frac{c' \bar{x}_{h} - \bar{\mu}_{h}}{\sqrt{c'(pz^{rs})c}} \to N(0,1)
\]
The
\[
\left\{ \frac{c' \bar{x}_{h} - \bar{\mu}_{h}}{\sqrt{c'(pz^{rs})c}} \right\}_{h=1}^{k}
\]
are independent and so it follows that
\[
\sum_{h=1}^{k} \frac{(c' \bar{x}_{h} - \bar{\mu}_{h})^2}{\sqrt{c'(pz^{rs})c}} \to \chi_{k}^2.
\]
We therefore approximate \( \xi_{p,k,\alpha} \) by \( \chi_{k,\alpha}^2 \).
In general linear combinations of the form \( \sum_{h=1}^{k} d_h c' \mu_h \) make sense only if all the components of \( \mu_h \) are measured in the same units. However even when they are not we may choose 
\[ c' = (\delta_{i1}, \ldots, \delta_{ip}) \] where \( \delta_{ij} \) is Kronecker's delta. In this case we would be looking at the ith component across the k populations.
APPENDIX I. COMPUTER PROGRAMS AND SAMPLE OUTPUT

Program I

C THIS PROGRAM IS DESIGNED TO CALCULATE THE MONTE CARLO
C THE CRITICAL-POINTS OF THE DISTRIBUTION OF THE
C TEST-STATISTIC TAGA FOR IJ = 5,10,15,25 AND
C K = 2,3,4,5,6. THE POWER AT THESE POINTS IS ALSO
C COMPUTED FOR DELTA-STAR VALUES .1,1.5 AND .2
C 2 VALUES .3, .4, .5, .6, .7, .8, .9, .1, .2, .3, .4, .5, .6
C
C DIMENSION C(50),Z(I1),H(4,6)
C DIMENSION A(4,5,500),S(5,4),H(11,5,4)
C DIMENSION XT(I6),QSTAR(I),CC(30),CD(80)
C DIMENSION A(11,6,4),S(11,5,4),M(4)
C DIMENSION FT(4,5,4,11),IO(4)
C REAL*2 ICONTU(4,5,4,11,80)/794000*0/
C REAL N0
C INTEGER 1F
C REAL N0
C 1FLAG=0
C N0 300 L=1,30
C (L)=1*(L-1)*(.25)
C 303 CONTINUE
C READ(I5,200)(Z(J),J=1,11)
C 200 FORMAT(I1F4.2)
C READ(I5,201)(QSTAR(I),I=1,3)
C 201 FORMAT(3F2.1)
C READ(I5,202)((MU(IJ,K),K=1,4),IJ=1,4)
C 202 FORMAT(5F9.7)
C READ(I5,204)((Q1(I),I=1,4)
C 204 FORMAT(4I2)
C CALL LLRAND
C M=1
C IX=524283
C
C THIS BEGINS THE MONTE CARLO.
C
C DO 1000 KJ=1,1000
C
C THE NEXT 5 STATEMENTS SET UP THE INITIAL
C OBSERVATIONS FOR 6 POPULATIONS WITH 4 DIFFERENT
C DELTA-STAR VALUES.
C
C 500 DO 2 K=1,6
C 2 M=1,25
C CALL NORMAL(I9,S,M)
C 1 IJ=1,4
C XI(J,K,4)=S=4U(I,J,K)
C 1 CONTINUE
C 3 CONTINUE
C 2 CONTINUE

254

DO 4 I=1,4
M=NI(I)
DO 5 K=1,6
XBAR=0
XX=0
DO 6 NM=1,10
XBAR=XBAR+X(I,K,NM)/NM
XX=XX+X(I,K,NM)**2
CONTINUE
SU(K,I)=(XX-NM*(XBAR**2))/((NM-1)
T=SU(K,I)*100.
IF(T.LE.-.99) GO TO 700
IFLAG=IFLAG+1
GO TO 500
700 DO 7 J=1,11
NP(K)=N(11,K,1)
DO 9 I=2,4
IF(N(11,K,I).GT.NP(K))NP(K)=N(11,K,I)
CONTINUE
KS=NP(K)-25
DO 10 L=1,KS
II=II+25
CALL NORMAL(IX,S,Y)
10 CONTINUE
1 CONTINUE
THE NEXT 11 STATEMENTS GENERATE THE SECOND-STAGE DATA FOR EACH POPULATION AND EACH DELTA-STAR.

DO 3 K=1,6
NP(K)=N(11,K,1)
DO 9 I=2,4
IF(N(11,K,I).GT.NP(K))NP(K)=N(11,K,I)
CONTINUE
KS=NP(K)-25
DO 10 L=1,KS
II=II+25
CALL NORMAL(IX,S,Y)
10 CONTINUE
11 CONTINUE
THE NEXT 27 STATEMENTS CALCULATE FTILDA FOR EACH COMBINATION OF NO, NUMBER-OF-POPULATIONS (2, 3, 4, 5, 6), Z, AND DELTA-STAR.

C

DO 12 IJ=1,4
DO 13 KK=1,5
XP=KK+1
DO 14 I=1,4
NM=NM(I)
DO 15 J=1,11
FT(IJ, KK, I, J)=0
TP=0
DO 16 K=1,XP
XA=0
XP=0
DO 17 NN=1,NN
XA=XA+X(IJ, K, NN)*A(J, K, I)
CONTINUE
KS=NN(J, K, I)-NN
DO 18 NN=1,KS
II=NN+NN
X3=X3+X(IJ, K, II)*B(J, K, I)
CONTINUE
XT(K)=XA+X3
TP=TP+XT(K)/XP
CONTINUE
DO 19 K=1,XP
FT(IJ, KK, I, J)=FT(IJ, KK, I, J)+
* ((XT(K)-TP)**2)/Z(J)
CONTINUE
XZ= FT(IJ, KK, I, J)
DO 20 L=1,SO

C THIS IS THE COUNTER WHICH COUNTS THE NUMBER OF TIMES FTILDA EXCEEDS C(L) FOR THE DIFFERENT COMBINATIONS OF NO, NUMBER-OF-POPULATIONS, Z, AND DELTA-STAR.

IF (XZ.GT.C(L))ICOUNT(IJ, KK, I, J, L)=
* ICOUNT(IJ, KK, I, J, L)+1
CONTINUE
CONTINUE
CONTINUE
CONTINUE
CONTINUE
CONTINUE
1000 CONTINUE
C END OF MONTE CARLO 10,000 REPETITIONS.
C THE REMAINDER OF THE PROGRAM PRINTS OUT THE RESULTS.
C
WRITE(6,101) IFLAG
101 FORMAT(*, 'IFLAG = ', IA)
   DO 21 IJ=2,4
      JK=IJ-1
      DO 22 KK=1,5
         K=KK+1
         DO 23 I=1,4
            IF (IJ.EQ.2.AND.KX.EQ.1.AND.I.EQ.1) WRITE(6,132)
            IF (IJ.EQ.2.AND.KX.EQ.1.AND.I.EQ.1) GO TO 701
22    FORMAT(*',///////////,41X,'TABLE 3.1'///)
23    WRITE(6,112)
132   FORMAT(*', //////////,38X,'TABLE 3.1 (CONT.)' ///)
701   WRITE(6,113) DSTAP(JK),K,NO(I)
113   FORMAT(*', 20X,'USTAR = ', F3.1)
      FORMAT(*', 13X,'K = ', I1, 13X,'NO = ', I2//)
      WRITE(6,110)
110   FORMAT(1X,'44X,'C'//)
      MLow=-7
      DO 25 I=1,10
         MLow=MLow+8
         MHIGH=MLow+7
         IF (MLow.GT.10) GO TO 500
      WRITE(6,109) (C (L ) , L = MLow,MHIGH)
25    FORMAT(*',15X,'Z',2X,2(2X,F5.2))
20    GO TO 601
600   WRITE(6,125)
125   FORMAT(*', //////////,38X,'TABLE 3.1 (CONT.)' ///)
      WRITE(6,113) DSTAP(JK),K,NO(I)
      WRITE(6,110)
      WRITE(6,105) (C(L),L=MLow,MHIGH)
105   FORMAT(1X,15X,'Z',2X,2(2X,F5.2))
601   DU 26 J=1,11
      WRITE(6,106) Z (J)
106   FORMAT(*',13X,F5.2)
      DO 27 L=MLow,MHIGH
         CC(L)=ICOUNT(IJ,KK,I,J,L)*(.0001)
         CC(L)=ICOUNT(IJ,KK,I,J,L)*(.0001)
      27 CONTINUE
      WRITE(6,137) (CD(L),L=MLow,MHIGH)
137   FORMAT(*',18X,3(3X,F4.3))
      WRITE(6,105) (CC(L),L=MLow,MHIGH)
105   FORMAT(*',10X,3(3X,F4.3))
      WRITE(7,130) (CD(L),L=MLow,MHIGH)
130   FORMAT(*', (CC(LL),LL=MLow,MHIGH)
<table>
<thead>
<tr>
<th>( Z )</th>
<th>1.00</th>
<th>1.25</th>
<th>1.50</th>
<th>1.75</th>
<th>2.00</th>
<th>2.25</th>
<th>2.50</th>
<th>2.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.408</td>
<td>0.355</td>
<td>0.311</td>
<td>0.278</td>
<td>0.253</td>
<td>0.230</td>
<td>0.210</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>0.422</td>
<td>0.371</td>
<td>0.320</td>
<td>0.284</td>
<td>0.267</td>
<td>0.240</td>
<td>0.219</td>
<td>0.200</td>
</tr>
<tr>
<td>0.80</td>
<td>0.408</td>
<td>0.355</td>
<td>0.311</td>
<td>0.278</td>
<td>0.253</td>
<td>0.230</td>
<td>0.210</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>0.428</td>
<td>0.377</td>
<td>0.332</td>
<td>0.299</td>
<td>0.271</td>
<td>0.244</td>
<td>0.220</td>
<td>0.200</td>
</tr>
<tr>
<td>0.60</td>
<td>0.408</td>
<td>0.355</td>
<td>0.311</td>
<td>0.278</td>
<td>0.253</td>
<td>0.230</td>
<td>0.210</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>0.434</td>
<td>0.383</td>
<td>0.343</td>
<td>0.308</td>
<td>0.277</td>
<td>0.249</td>
<td>0.226</td>
<td>0.200</td>
</tr>
<tr>
<td>0.40</td>
<td>0.408</td>
<td>0.355</td>
<td>0.311</td>
<td>0.278</td>
<td>0.253</td>
<td>0.230</td>
<td>0.210</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>0.449</td>
<td>0.399</td>
<td>0.355</td>
<td>0.321</td>
<td>0.291</td>
<td>0.265</td>
<td>0.241</td>
<td>0.219</td>
</tr>
<tr>
<td>0.20</td>
<td>0.408</td>
<td>0.355</td>
<td>0.311</td>
<td>0.278</td>
<td>0.253</td>
<td>0.230</td>
<td>0.210</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>0.489</td>
<td>0.441</td>
<td>0.369</td>
<td>0.359</td>
<td>0.328</td>
<td>0.299</td>
<td>0.274</td>
<td>0.251</td>
</tr>
<tr>
<td>0.10</td>
<td>0.408</td>
<td>0.355</td>
<td>0.311</td>
<td>0.278</td>
<td>0.253</td>
<td>0.230</td>
<td>0.210</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>0.561</td>
<td>0.508</td>
<td>0.472</td>
<td>0.436</td>
<td>0.405</td>
<td>0.374</td>
<td>0.347</td>
<td>0.328</td>
</tr>
<tr>
<td>0.08</td>
<td>0.408</td>
<td>0.355</td>
<td>0.311</td>
<td>0.278</td>
<td>0.253</td>
<td>0.230</td>
<td>0.210</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>0.594</td>
<td>0.546</td>
<td>0.501</td>
<td>0.463</td>
<td>0.428</td>
<td>0.400</td>
<td>0.372</td>
<td>0.347</td>
</tr>
<tr>
<td>0.06</td>
<td>0.408</td>
<td>0.355</td>
<td>0.311</td>
<td>0.278</td>
<td>0.253</td>
<td>0.230</td>
<td>0.210</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>0.639</td>
<td>0.597</td>
<td>0.554</td>
<td>0.515</td>
<td>0.482</td>
<td>0.452</td>
<td>0.423</td>
<td>0.396</td>
</tr>
<tr>
<td>0.04</td>
<td>0.408</td>
<td>0.355</td>
<td>0.311</td>
<td>0.278</td>
<td>0.253</td>
<td>0.230</td>
<td>0.210</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>0.728</td>
<td>0.687</td>
<td>0.649</td>
<td>0.612</td>
<td>0.576</td>
<td>0.549</td>
<td>0.519</td>
<td>0.489</td>
</tr>
<tr>
<td>0.02</td>
<td>0.408</td>
<td>0.355</td>
<td>0.311</td>
<td>0.278</td>
<td>0.253</td>
<td>0.230</td>
<td>0.210</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>0.657</td>
<td>0.833</td>
<td>0.809</td>
<td>0.766</td>
<td>0.734</td>
<td>0.704</td>
<td>0.674</td>
<td>0.646</td>
</tr>
<tr>
<td>0.01</td>
<td>0.408</td>
<td>0.355</td>
<td>0.311</td>
<td>0.278</td>
<td>0.253</td>
<td>0.230</td>
<td>0.210</td>
<td>0.192</td>
</tr>
<tr>
<td></td>
<td>0.952</td>
<td>0.944</td>
<td>0.934</td>
<td>0.924</td>
<td>0.915</td>
<td>0.906</td>
<td>0.896</td>
<td>0.887</td>
</tr>
</tbody>
</table>
PROGRAM II

C THIS PROGRAM IS DESIGNED TO COMPARE THE
C NOMINAL SIZES OF THE USUAL F TEST AND
C FTILDA AND ITS APPROXIMATIONS FOR VARIOUS
C CONFIGURATIONS OF STANDARD DEVIATIONS AND
C SAMPLE SIZES. IT IS ALSO USED TO COMPARE
C THE POWER OF THESE STATISTICS FOR THE
C NOMINAL SIZES WITH DELTA-STAR VALUES .1,
C AND 5.
C
DIMENSION NN(5,4,6),X(6,500),Y(6,500)
DIMENSION FC(5,4,3),FTC(5,3),FTCLA(5,3)
DIMENSION FTC(5,3),XBAR(6),XXO(6),XX(6)
DIMENSION XBAR(6),S0(6),S(6),A(6),B(6)
DIMENSION NM(6),XT(6)
INTEGER STD(5,6,6)
REAL MU(6)
REAL EN(5,4,6)/120*0.0/
REAL ICF(5,4,6,3)/360*0.0/
REAL ICFT(5,4,6,3)/360*0.0/
REAL ICFTLA(5,4,6,3)/360*0.0/
REAL ICFTAL(5,4,6,3)/360*0.0/
IFLAG=0
ZYZ=(9.)/(7.0)
DSTAR=.1

C THE NEXT 22 STATEMENTS READ IN THE VALUES
C OF THE SAMPLE SIZES, THE STANDARD DEVIATIONS,
C THE CRITICAL POINTS OF THE DISTRIBUTIONS
C OF F AND FTILDA AND THE APPROXIMATE CRITICAL
C POINTS FOR FTILDA AND THE POPULATION MEANS.
C
DO 1 I=1,5
DO 1 J=1,4
READ(5,101)(NN(I,J,K),K=1,6)
101 FORMAT(6I2)
1 CONTINUE
DO 2 I=1,5
DO 2 J=1,6
READ(5,102)(STD(I,J,K),K=1,5)
102 FORMAT(6F6.4)
2 CONTINUE
DO 3 I=1,5
DO 3 J=1,4
READ(5,103)(FC(I,J,K),K=1,3)
103 FORMAT(3F6.4)
3 CONTINUE
DO 4 I=1,5
READ(5,103)(FTC(I,J),J=1,3)
READ(5,103)(FTCLA(I,J),J=1,3)
READ(5,103)(FTCLA(I,J),J=1,3)
4 CONTINUE
READ(5,203)(MU(I),I=1,6)
203 FORMAT(6F9.7)
DO 40 I=1,5
DO 41 J=1,2
FTCLA(I,J)=FTCLA(I,J)**2
41 CONTINUE
40 CONTINUE
DO 50 I=1,5
DO 51 J=1,3
FTCLA(I,J)=FTCLA(I,J)**3
51 CONTINUE
50 CONTINUE
CALL LLRAND
MMM=1
IX=524289
C
C THIS BEGINS THE MONTE CARLO EXPERIMENT.
C
DO 1000 KJ=1,10000
C
THE LOOP FROM STATEMENT 600 TO STATEMENT
C 6 GENERATE THE STATISTICS F AND FTILDA
C AND COMPARE THEM TO THEIR CRITICAL POINTS.
C
DO 5 X=1,6
DO 5 N=1,20
CALL NORMAL(IX,SS,MMM)
X(K,N)=SS
5 CONTINUE
DO 6 I=1,5
K=I+1
600 DO 6 J=1,6
DO 7 KK=1,K
DO 7 N=1,20
Y(KK,N)=STD(I,J,KK)*X(KK,N)+MU(KK)
7 CONTINUE
DO 6 NS=1,4
X8=0
MM=0
WW=0
DO 8 KI=1,K
X3ARJ(KI)=U
X3AR(KI)=0
XXU(KI)=0
XX(KI) = 0
N = N(I, NS, KI)
MM = MM + N
WW = WW + STD(I, J, KI)**2
DO 10 JJ = 1, N
X3(1, KI) = XBAR(KI) + Y(KI, JJ)/N
XX(KI) = XX(KI) + Y(KI, JJ)**2
XBAR(KI) = XBAR(KI) + Y(KI, JJ)/N
CONTINUE
DO 9JJ = 1, 10
XX(JI, KI) = XXO(KI) - Y(KI, JJ)**2
XBARO(KI) = XBARO(KI) + Y(KI, JJ)/10
CONTINUE
S(JI) = (XX(JI) - N*(XBARO(JI)**2))/(N - 1)
S0(JK) = (XXO(KI) - 10*(XBARO(KI)**2))/9
CONTINUE
Z = AA/MM
XX = XX/MM
YY = 0
ZZ = 0
DO 11 JL = 1, K
YY = YY + IN(I, NS, JL)*(XBAR(JL) - XB)**2)/(K - 1)
ZZ = ZZ + ((NN(I, NS, JL) - 1)*S(JL))/(MM-K)
YZ = S(JL)/ZO
IY = IFIX(YZ)
IZ = IY + 1
IF(IY <= 500) IFLAG = IFLAG + 1
IF(IY <= 500) GO TO 660
NR = U
NM(JL) = MAXO(MR, IY)
EN(I, NS, JL) = EN(I, NS, JL) + NM(JL)*(.0001)
B(JL) = (1./NM(JL))*((10*NM(JL)*ZO - S(JL))/
(*((NM(JL) - 10)*S(JL))**(.5))
A(JL) = (1 - (NM(JL) - 10)*B(JL))/10
CONTINUE
C F IS THE VALUE OF THE F STATISTIC.
C
F = YY/ZZ
DO 12 JL = 1, K
KS = NM(JL) - 20
IF(KS <= 0) GO TO 12
DO 13 L = 1, KS
MN = ZO + L
CALL NORMAL(IJ, SS, MN)
Y(JL, MN) = STD(I, J, JL)*(SS + MU(JL))
CONTINUE
CONTINUE
CONTINUE
TP = 0
DO 14 JL = 1, K
XA = 0
XB = 0
DO 15 N = 1, 10
XA = XA + A(JL) * Y(JL, N)
CONTINUE
KS = NM(JL) - 10
DO 16 N = 1, KS
MN = 10 + N
XB = XB + B(JL) * Y(JL, MN)
CONTINUE
XT(JL) = XA + XB
TP = TP + XT(JL) / K
CONTINUE

C THE NEXT LOOP GENERATES THE VALUE FT OF C THE STATISTIC FTILDA.
C
DO 17 JL = 1, K
FT = FT + ((XT(JL) - TP) ** 2) / Z0
CONTINUE
C THE NEXT LOOP COMPARES THE VALUES OF C F AND FTILDA AGAINST THEIR RESPECTIVE C CRITICAL VALUES.
C
DO 18 L = 1, 3
IF (F .GT. FC(I, NS, J, L)) ICF(I, NS, J, L) =
* ICF(I, NS, J, L) + .0001
IF (FT .GT. FTC(I, L)) ICFT(I, NS, J, L) =
* ICFT(I, NS, J, L) + .0001
IF (FT .GT. FTCLA(I, L)) ICFTLA(I, NS, J, L) =
* ICFTLA(I, NS, J, L) + .0001
IF (FT .GT. FTCA(I, L)) ICFTA(I, NS, J, L) =
* ICFTA(I, NS, J, L) + .0001
CONTINUE
6 CONTINUE
1000 CONTINUE
C
C THIS ENDS THE 'MONTE CARLO EXPERIMENT'.
C THE REMAINDER OF THE PROGRAM PRINTS C OUT THE RESULTS.
C
WRITE(6, 230) IFLAG
230 FORMAT(IX, 'IFLAG=', I5)
DO 20 I = 1, 5
K = I + 1
M = I + 1
DO 21 NS = 1, 4
    IF (K .EQ. 2 .AND. NS .EQ. 1) WRITE (6, 701)
    FORMAT ('1', '///', '43X', 'TABLE 3.3' ///)
    IF (K .EQ. 2 .AND. NS .EQ. 1) GO TO 801
    WRITE (6, 702)
    FORMAT ('1', '///', '37X', 'TABLE 3.3 (CONT.)' ///)
701   WRITE (6, 703) K
    FORMAT ('+', '45X', 'K = ', ', ', 'I2' ///)
    WRITE (6, 733) DSTAR
    IF (K .EQ. 2 .AND. NS .EQ. 1) GO TO
    WRITE (6, 704) K
    FORMAT ('+', '42X', 'DSTAR = ', ', ', 'F2.1' ///)
    IF (K .EQ. 1) WRITE (6, 705) (NN (I, NS, KK), KK = 1, M)
    IF (I .EQ. 2) WRITE (6, 706) (NN (I, NS, KK), KK = 1, M)
    IF (I .EQ. 3) WRITE (6, 707) (NN (I, NS, KK), KK = 1, M)
    IF (I .EQ. 4) WRITE (6, 708) (NN (I, NS, KK), KK = 1, M)
704   FORMAT ('+', '37X', 'SAMPLE SIZES', '2(1X, I2)' ///)
705   FORMAT ('+', '35X', 'SAMPLE SIZES', '3(1X, I2)' ///)
706   FORMAT ('+', '33X', 'SAMPLE SIZES', '4(1X, I2)' ///)
707   FORMAT ('+', '31X', 'SAMPLE SIZES', '5(1X, I2)' ///)
708   FORMAT ('+', '29X', 'SAMPLE SIZES', '6(1X, I2)' ///)
711   FORMAT ('+', '16X', 'STANDARD', '14X', 'F', '20X', 'FT', '10X', 'E(N)' ///)
    WRITE (6, 712)
    IF (K .EQ. 1) WRITE (6, 713) (NN (I, NS, J, L), L = 1, 3)
    IF (I .EQ. 2) WRITE (6, 714) (STD (I, J, KK), KK = 1, M)
DO 22 J = 1, 6
    WRITE (6, 715) (ICF (I, NS, J + L), L = 1, 3)
    WRITE (6, 716) (ICFT (I, NS, J, L), L = 1, 3)
22 CONTINUE
    WRITE (6, 717) EM (I, NS, J)
    FORMAT ('+', '72X', 'F4.0' ///)
718   FORMAT ('J', '16X', 'STANDARD', '13X', 'FTLA', '18X', 'FTA ///)
    WRITE (6, 718)
    WRITE (6, 719)
719   FORMAT ('+', '31X', '10%', '4X', '5%', '4X', '1%' ///)
DO 23 J = 1, 6
WRITE(6,721)(STD(I,J,KK),KK=1,M)
721 FORMAT('**',16X,6(I1,1X))
WRITE(6,715)(ICFLA(I,NS,J,L),L=1,3)
WRITE(6,720)(ICFTA(I,NS,J,L),L=1,2)
720 FORMAT('**',55X,F4.3,3X,F4.3//)
23 CONTINUE
21 CONTINUE
20 CONTINUE
STOP
END

//GO SYSIN DD *
6 6
612
1212
1220
6 6 6
61215
111111
81520
6 6 6 6
6 6 810
12121212
6101620
6 6 6 6 6
6 8 81212
1313131313
610161820
6 6 6 6 6 5
111111111111
4 610121520
6 610102020
11
22
12
21
13
31
111
123
321
113
311
222
1111
11113
3111
1223
3221
267
### Table 3.2 (Cont.)

*K = 3*

**Sample Sizes 6 6 6**

<table>
<thead>
<tr>
<th>Standard Deviations</th>
<th>F</th>
<th>N</th>
<th>T</th>
<th>FTA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
</tr>
<tr>
<td>1 1 1</td>
<td>0.096</td>
<td>0.048</td>
<td>0.010</td>
<td>0.104</td>
</tr>
<tr>
<td>1 2 3</td>
<td>0.116</td>
<td>0.066</td>
<td>0.018</td>
<td>0.105</td>
</tr>
<tr>
<td>3 2 1</td>
<td>0.112</td>
<td>0.067</td>
<td>0.019</td>
<td>0.105</td>
</tr>
<tr>
<td>1 1 3</td>
<td>0.137</td>
<td>0.082</td>
<td>0.029</td>
<td>0.106</td>
</tr>
<tr>
<td>3 1 1</td>
<td>0.132</td>
<td>0.085</td>
<td>0.029</td>
<td>0.104</td>
</tr>
<tr>
<td>2 2 2</td>
<td>0.096</td>
<td>0.048</td>
<td>0.010</td>
<td>0.104</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard Deviations</th>
<th>FTA</th>
<th>N</th>
<th>T</th>
<th>FTA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
</tr>
<tr>
<td>1 1 1</td>
<td>0.103</td>
<td>0.058</td>
<td>0.016</td>
<td>0.103</td>
</tr>
<tr>
<td>1 2 3</td>
<td>0.103</td>
<td>0.058</td>
<td>0.017</td>
<td>0.103</td>
</tr>
<tr>
<td>3 2 1</td>
<td>0.103</td>
<td>0.056</td>
<td>0.016</td>
<td>0.104</td>
</tr>
<tr>
<td>1 1 3</td>
<td>0.104</td>
<td>0.056</td>
<td>0.016</td>
<td>0.104</td>
</tr>
<tr>
<td>3 1 1</td>
<td>0.102</td>
<td>0.055</td>
<td>0.018</td>
<td>0.104</td>
</tr>
<tr>
<td>2 2 2</td>
<td>0.103</td>
<td>0.058</td>
<td>0.016</td>
<td>0.103</td>
</tr>
</tbody>
</table>
REFERENCES


269


Rinott, Y. (1975): "On two-stage procedures for selecting the population with the largest mean from several normal populations with unknown variances," "unpublished paper."


Tamhane, A.C. (1975): "A 2- and 3-stage screening procedure for selecting the population having the largest mean from k normal populations with a common unknown variance," unpublished paper.


Welch, B.L. (1947): "The generalization of "Student's" problem when several different population variances are involved," Biometrika, Vol. 34, pp. 28-35.

