Selected topics in simultaneous statistical inference

Dorothy Anne Vanderburg

The University of Montana

Follow this and additional works at: https://scholarworks.umt.edu/etd

Let us know how access to this document benefits you.

Recommended Citation
https://scholarworks.umt.edu/etd/8334

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.
SELECTED TOPICS IN SIMULTANEOUS STATISTICAL INference

By

Dorothy Anne Vanderburg

B.A. University of North Carolina at Greensboro, May 31, 1964

for the degree of

Master of Arts

UNIVERSITY OF MONTANA

1970

Approved by:

Chairman, Board of Examiners

Dean, Graduate School

Date
ACKNOWLEDGEMENTS

I gratefully acknowledge the help of Dr. Don Loftsgaarden in all stages of the development of this thesis. Without his help I could not have mastered the concepts involved. His suggestions and corrections during the writing of the thesis were invaluable. I am especially grateful for his patience and encouragement at the times I needed them most.

D. A. V.
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>I. A GENERAL INTRODUCTION TO SIMULTANEOUS STATISTICAL INFERENCE</td>
<td>2</td>
</tr>
<tr>
<td>II. TUKEY'S STUDENTIZED RANGE</td>
<td>8</td>
</tr>
<tr>
<td>2.1 Introduction and derivation</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Contrasts</td>
<td>10</td>
</tr>
<tr>
<td>2.3 The studentized augmented range random variable and simultaneous confidence intervals for all linear combinations of means</td>
<td>12</td>
</tr>
<tr>
<td>2.4 Generalization of Tukey's method to dependent random variables</td>
<td>14</td>
</tr>
<tr>
<td>2.5 Applications of the studentized range technique</td>
<td>15</td>
</tr>
<tr>
<td>III. SCHEFFE' F PROJECTIONS</td>
<td>19</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>19</td>
</tr>
<tr>
<td>3.2 Extension of the Scheffé method to dependent random variables</td>
<td>22</td>
</tr>
<tr>
<td>3.3 Scheffé F projections and hypothesis testing</td>
<td>22</td>
</tr>
<tr>
<td>3.4 Applications</td>
<td>25</td>
</tr>
<tr>
<td>3.4.1 Application to one-way design</td>
<td>25</td>
</tr>
<tr>
<td>3.4.2 Application to two-way classification</td>
<td>27</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>3.4.3 Application to regression: Simple linear regression and general regression.</td>
<td>29</td>
</tr>
<tr>
<td>3.4.4 Application to prediction in regression analysis</td>
<td>31</td>
</tr>
<tr>
<td>IV. BONFERRONI T STATISTICS AND STUDENTIZED MAXIMUM MODULUS TECHNIQUES</td>
<td>33</td>
</tr>
<tr>
<td>4.1 Bonferroni t statistics</td>
<td>33</td>
</tr>
<tr>
<td>4.2 Studentized maximum modulus technique</td>
<td>37</td>
</tr>
<tr>
<td>4.3 Multiple range tests</td>
<td>38</td>
</tr>
<tr>
<td>V. NUMERICAL EXAMPLES</td>
<td>40</td>
</tr>
<tr>
<td>5.1 Application of the Scheffé, Tukey, and Bonferroni techniques to randomized complete block design</td>
<td>40</td>
</tr>
<tr>
<td>5.2 Application of the Scheffé, Tukey, and Bonferroni techniques to a two-way classification design</td>
<td>46</td>
</tr>
<tr>
<td>5.3 Application of the Scheffé, Tukey, and Bonferroni techniques to a one-way classification design</td>
<td>50</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>54</td>
</tr>
</tbody>
</table>
INTRODUCTION

Often an experimenter conducts a random experiment to collect data to help him make statements about certain unknown constants called parameters. Because the results of a random experiment vary, the experimenter can never be one hundred percent sure that a particular statement he makes is correct. However, in many cases, there are available to him tools (techniques of statistical inference), which assure him a high degree of probability that the statement he makes will be correct. Of course, once a particular statement is made it is either true or it is false.

Many of the procedures of statistical inference can be used for only one statement. More often than not, the experimenter will be concerned with making more than one statement. Of course it would be desirable for him to be assured a high degree of probability that all his statements simultaneously will be correct. To achieve this goal, he must use one of the techniques of simultaneous statistical inference.

This thesis is concerned with the theory and application of three major techniques of simultaneous statistical inference: Scheffé F projections, Tukey's Studentized range and Bonferroni T statistics. The last chapter is composed of numerical examples illustrating the three techniques.
CHAPTER I
A GENERAL INTRODUCTION TO SIMULTANEOUS
STATISTICAL INFERENCE

In any given random experiment, the experimenter is concerned with making statements which accurately and adequately describe the results of the experiment. These statements are often of the form of confidence intervals for the parameters, or the statements may concern hypotheses about parameters. He is interested in saying something about the correctness of these statements.

There are two methods of approach. First, if $S_i$, $i = 1, 2, \ldots, n$ are statements made on the basis of the random experiment and $A_i$, $i = 1, 2, \ldots, n$ are the events that the statements $S_i$ are correct, he may deal with the statements individually and determine $\alpha_i = p(A_i)$, $i = 1, 2, \ldots, n$, $(1 - \alpha_i = p(A_i))$. However, it seems likely that in many cases the experimenter would be interested in knowing the probability that all the statements, $S_i$, $i = 1, 2, \ldots, n$, simultaneously are correct; i.e., $p(\bigcap_{i=1}^{n} A_i)$ is of concern. This is the subject matter of simultaneous statistical inference.

Suppose, for example, that two means, $\mu_1$ and $\mu_2$, are under consideration. Consider statements $S_1: L_1 \leq \mu_1 \leq L_1'$ and $S_2: L_2 \leq \mu_2 \leq L_2'$, where $L_1, L_1', L_2, L_2'$ are random variables. Let $A_1$ be the event that $S_1$ is true and $A_2$
be the event that $S_2$ is correct. Suppose that $1 - \alpha_1 = p(A_1) = p(L_1 \leq \mu_1 \leq L_1') = .95$ and that $1 - \alpha_2 = p(A_2) = p(L_2 \leq \mu_2 \leq L_2') = .95$. In other words, both intervals are 95 percent confidence intervals.

Many times, however, he will want to know the probability that both $S_1$ and $S_2$ are correct simultaneously; that is, he will want $p(A_1 \cap A_2) = p(L_1 \leq \mu_1 \leq L_1' \text{ and } L_2 \leq \mu_2 \leq L_2')$. Quite often because of distributional difficulties that arise, this probability is difficult to determine exactly, but knowing $\alpha_1$ and $\alpha_2$, it is always possible to get a bound for this probability. This bound is obtained from the following argument:

$$p(A_1 \cup A_2) \leq p(A_1) + p(A_2) = \alpha_1 + \alpha_2.$$  Hence, $p(A_1 \cap A_2) = p(A_1 \cup A_2) \geq 1 - \alpha_1 - \alpha_2$.

Thus, in the above example, the experimenter could conclude that $p(L_1 \leq \mu_1 \leq L_1' \text{ and } L_2 \leq \mu_2 \leq L_2') \geq 1 - .05 - .05 = .90$. If he desires this probability to be 95 percent instead, he might decide to determine random variables $L_3$, $L_3'$, $L_4$, $L_4'$ with the property that $p(L_3 \leq \mu_1 \leq L_3') = .025$ and $p(L_4 \leq \mu_2 \leq L_4') = .025$. Then, as above, it is true that $p(L_3 \leq \mu_1 \leq L_3' \text{ and } L_4 \leq \mu_2 \leq L_4') = .95$. Much of simultaneous statistical inference is concerned with making exact statements about $p(A_1 \cap A_2)$ rather than using bounds as illustrated above.

A situation which is easily handled but which seldomly occurs in actual practice is the one in which $S_1$ and $S_2$ are
independent. In this case, \( p(A_1 \cap A_2) = p(L_1 \leq \mu_1 \leq L_1' \quad \text{and} \quad L_2 \leq \mu_2 \leq L_2') = p(A_1) \cdot p(A_2) = (1 - \alpha_1) \cdot (1 - \alpha_2) \). With \( \alpha_1 = \alpha_2 = .05 \), this probability turns out to be \(.9025\) which differs only slightly from the bound obtained above.

Another illustration of simultaneous inference can be taken from hypothesis testing. If the parameters under consideration are \( \Theta_1, \Theta_2, \ldots, \Theta_s \), the experimenter may choose to consider separate hypotheses \( H_{oi} : \Theta_i = \Theta_{io} \), where \( \Theta_{io} \) is some specified constant, \( i = 1, 2, \ldots, s \). On the other hand, he may be interested in testing a single hypothesis \( H_o : \Theta_1 = \Theta_{10}, \Theta_2 = \Theta_{20}, \ldots, \Theta_s = \Theta_{so} \), which involves all the parameters simultaneously.

Consider the case of two parameters \( \Theta_1 \) and \( \Theta_2 \). The hypotheses \( H_{o1} : \Theta_1 = \Theta_{10} \), where \( \Theta_{10} \) is some real constant and \( H_{o2} : \Theta_2 = \Theta_{20} \), where \( \Theta_{20} \) is some real constant, are to be tested. When \( H_{oi} \) is true then statement \( S_i : \Theta_i = \Theta_{io} \) is the correct statement to make. If \( A_i \) is the event that \( S_i \) is made when \( H_{oi} \) is true, then \( 1 - \alpha_i = p(A_i) = p(\text{accept } H_{oi} | H_{oi} \text{ is true}) \). If \( \alpha_1 = \alpha_2 = .05 \), then \( p(A_1) = .95 \) and \( p(A_2) = .95 \). However for various reasons it may be more desirable to test a single hypothesis \( H_o \) which involves \( \Theta_1 \) and \( \Theta_2 \) simultaneously. If \( H_o : \Theta_{10}, \Theta_2 = \Theta_{20} \), with \( \Theta_{10} \) and \( \Theta_{20} \) as before, the statement, \( S : \Theta_1 = \Theta_{10}, \Theta_2 = \Theta_{20} \), is the correct statement to make when \( H_o \) is true. If individual tests for \( \Theta_1 \) and \( \Theta_2 \) are used as they were earlier in this paragraph, then the following statements can be
made concerning $H_0$. If $A$ is the event that statement $S$ is made when $H_0$ is true, then $p(A) = p(A_1 \cap A_2) \geq 1 - \alpha_1 - \alpha_2$. If $\alpha_1 - \alpha_2 = .05$, then $P(A) = p(\text{accept } H_0 | H_0 \text{ is true}) \geq .90$.

So consider the matter more formally, let $\mathcal{F} = \{S_f\}$ be a family of statements and $N(\mathcal{F})$ be the number of statements in the family. If $N_w(\mathcal{F})$ is the number of incorrect statements in the family, then the error rate for the family is $Er(\mathcal{F}) = \frac{N_w(\mathcal{F})}{N(\mathcal{F})}$. The error rate is a random variable whose variable whose distribution depends on the procedure used in making the family of statements and on the underlying probability structure. Since the family error rate is a random variable, one concept that arises naturally is that of the probability of a non-zero error rate.

The probability error rate, denoted $P(\mathcal{F})$ is defined as $P(\mathcal{F}) = p\left[\frac{N_w(\mathcal{F})}{N(\mathcal{F})} > 0\right] = p\left[N_w(\mathcal{F}) > 0\right]$. The last equality follows since $N(\mathcal{F})$ is constant. If the family $\mathcal{F}$ contains an infinite number of statements we define $P(\mathcal{F}) = p\left[N_w(\mathcal{F}) > 0\right]$. Thus, the probability error rate is the probability that at least one of the statements in family $\mathcal{F}$ is incorrect. In the event that $\mathcal{F}$ contains only one statement, $S$, which is a confidence interval for some parameter, $1 - P(\mathcal{F})$ is the familiar $1 - \alpha$. Likewise, if $\mathcal{F}$ contains a single statement arising from testing a hypothesis about a parameter, then $P(\mathcal{F})$ is simply the probability of a type I error.
It seems obvious that a family with a large number of statements \( N(\mathcal{F}) \) will have a greater chance of having at least one incorrect statement than will a family with a smaller number of statements. Therefore, to achieve the same \( P(\mathcal{F}) \) the statements for the larger family must be weaker. Essentially this means that as the number of statements in a family increases, confidence intervals become wider and critical regions for tests become smaller if \( \alpha \) is held constant.

In the following chapters, some rather sophisticated techniques for simultaneous confidence inference will be dealt with in detail. It is worthwhile at this point, however, to consider a simple but often useful technique for simultaneous inference, the Bonferroni inequality. This technique has already been illustrated in earlier examples of this chapter. Suppose that \( \mathcal{F} \) contains only a finite number of statements. For each statement \( S_f \) in \( \mathcal{F} \), let \( A_f \) be the event that \( S_f \) is correct and let \( \alpha_f = p(\overline{A_f}) \). Then, \( P(\mathcal{F}) = p(\bigcup A_f) \leq \sum_f p(\overline{A_f}) = \sum \alpha_f \). Hence \( p(\text{all statements are correct}) = p(\bigcap A_f) = 1 - P(\mathcal{F}) \geq 1 - \sum \alpha_f \). The Bonferroni inequality, \( p(\bigcap A_f) \geq 1 - \sum \alpha_f \), gives a sometimes crude bound which relates the individual probabilities for the statements to \( P(\mathcal{F}) \).

If the statements in \( \mathcal{F} \) are confidence intervals for parameters, then the Bonferroni inequality provides a lower bound for the probability that all of the parameters lie
within their respective confidence intervals simultaneously. It is a nice tool especially when attempts to use other techniques lead to distributional difficulties. Furthermore, if the number of statements is not too large and the $\alpha_f$ are small, the bound is good and generally much simpler to use than any of the exact procedures.

The simplest case, which is seldomly encountered, is the one in which the statements are independent. The probability that a particular statement $S_f$ is correct is $1 - \alpha_f$. Using the independence of the statements, we can conclude that the probability that all the statements are correct is equal to the product of the probabilities that the individual statements are correct; that is, $1 - P(\mathcal{H}) = \pi_f(1 - \alpha_f)$. Here we obtain strict equality as opposed to the inequality we got by using the Bonferroni inequality.
CHAPTER II
TUKEY'S STUDENTIZED RANGE

Section 2.1 Introduction and derivation

The first technique for simultaneous inference to be considered is Tukey's studentized range. It can be used both for testing hypotheses and for obtaining simultaneous confidence intervals.

Let $X_1, X_2, \ldots, X_r$ be a random sample from a $N(0, 1)$ distribution and let $R = X(r) - X(1)$ be their range. If $W$ is distributed as a chi-square with $v$ degrees of freedom and is independent of the $X_i$, $i = 1, 2, \ldots, r$, then

$$Q = \frac{R}{\sqrt{\frac{W}{v}}}$$

is a studentized range random variable with $r$ and $v$ degrees of freedom.

Suppose $Y = (y_1, y_2, \ldots, y_r)$ has a multivariate normal distribution with mean $\mu = (\mu_1, \mu_2, \ldots, \mu_r)$, and covariance matrix $\sigma^2 I$ and that $\frac{vS^2}{\sigma^2}$ is an independent chi-square random variable with $v$ degrees of freedom. The statistician may be interested in testing $H_0: \mu_1 = \mu_2 = \ldots = \mu_r$ (or equivalently $H_0: \mu_i - \mu_{i'} = 0$ for $i \neq i'$); he may desire simultaneous confidence intervals for pairwise differences in means. In either case, the studentized range technique is embodied in the following probability statement:

$$(1) \quad p\{ |(y_i - y_{i'}) - (\mu_i - \mu_{i'})| \leq q_{r,v}^\alpha S, \ i, \ i' = 1, 2, \ldots, r, i \neq i' \} = 1 - \alpha.$$
where $q_{r,v}^\alpha$ is the upper 100 $\alpha$ percent point of the studentized range distribution with $r$ and $v$ degrees of freedom.

To verify this statement, let us first define $y_i^* = \frac{y_i - \mu_i}{\sigma}$, $i = 1, 2, \ldots, r$. Clearly $y_i^* \sim n(0, 1)$, $i = 1, 2, \ldots, r$. Now, consider

$$Q = \max_{i \neq i'} \left\{ \frac{|y_i - \mu_i - y_{i'} - \mu_{i'}|}{\sigma} \right\} = \max_{i \neq i'} \frac{|(y_i - \mu_i) - (y_{i'} - \mu_{i'})|}{S} \sqrt{\frac{v}{S^2}} \sigma^2 \cdot v$$

By definition $Q$ is a studentized range random variable with $r$ and $v$ degrees of freedom. Therefore,

$$p\{\max_{i \neq i'} \frac{|(y_i - \mu_i) - (y_{i'} - \mu_{i'})|}{S} \leq q_{r,v}^\alpha \} = 1 - \alpha.$$ 

However, $\max_{i \neq i'} \{|(y_i - \mu_i) - (y_{i'} - \mu_{i'})|\} \leq q_{r,v}^\alpha \cdot S$ if and only if $|(y_i - \mu_i) - (y_{i'} - \mu_{i'})| \leq q_{r,v}^\alpha \cdot S$ for all $i, i' = 1, 2, \ldots, r, i \neq i'$. Hence $p\{\max_{i \neq i'} \{|(y_i - \mu_i) - (y_{i'} - \mu_{i'})|\} \leq q_{r,v}^\alpha \cdot S\} = 1 - \alpha$ if and only if $p\{|(y_i - \mu_i) - (y_{i'} - \mu_{i'})| \leq q_{r,v}^\alpha \cdot S\}$, for all $i, i' = 1, 2, \ldots, r, i \neq i'$.

Thus we obtain a family $\mathcal{H}$ of $(\binom{r}{2})$ confidence statements which has a probability error rate equal to $\alpha$. The statements are of the form: $\mu_i - \mu_{i'} \in y_i - y_{i'} \pm q_{r,v}^\alpha \cdot S$, for all $i, i' = 1, 2, \ldots, r, i \neq i'$. The hypothesis $H_0: \mu_i - \mu_{i'} = 0, i \neq i'$, $i, i' = 1, 2, \ldots, r$ can be tested simply by checking the confidence intervals to determine whether zero is in each. If zero is in each confidence
interval accept $H_0$; otherwise, reject $H_0$.

Section 2.2 Contrasts

Concern with pairwise comparisons of means can be generalized to the study of contrasts. A contrast is a linear combination, \( \sum_{i=1}^{r} c_i \mu_i \), where \( \sum_{i=1}^{r} c_i = 0 \). The linear space of the totality of all contrasts will be denoted \( \mathcal{Z}_c \). A vector \( (c_1 c_2 \ldots c_r) \) in this space will be denoted \( c \).

The probability statement concerning the family of contrasts is

\[
(2) \quad P\{ \sum_{i=1}^{r} c_i \mu_i \in \sum_{i=1}^{r} c_i y_i + q \alpha, S \sum_{i=1}^{r} \frac{|c_i|}{2} \}, \text{ for all } c \in \mathcal{Z}_c
\]

\( = 1 - \alpha \), where \( y_1, y_2, \ldots, y_r \) and \( S \) are as previously defined. The following lemma is used to obtain statement (2). The general outline of its proof and of all other proofs in this paper is taken from Simultaneous Statistical Inference by Rupert P. Miller, Jr.

Lemma 1: \( |y_i - y_i'| \leq c \) for all \( i, i' = 1, 2, \ldots, r \), \( i \neq i' \) if and only if

\[
|\sum_{i=1}^{r} c_i y_i| \leq c \sum_{i=1}^{r} \frac{|c_i|}{2} \text{ for all } c \in \mathcal{Z}_c.
\]

(a) If \( |y_i - y_i'| \leq c \) for all \( i, i' = 1, 2, \ldots, r \), \( i \neq i' \)

then \( |\sum_{i=1}^{r} c_i y_i| \leq c \sum_{i=1}^{r} \frac{|c_i|}{2} \text{ for all } c \in \mathcal{Z}_c \).

Proof: If \( c_i = 0, i = 1, 2, \ldots, r \), the lemma is obviously true. Suppose \( c_i \neq 0, i = 1, 2, \ldots, r \). Let \( P = \{i | c_i > 0\} \) and \( N = \{i | c_i < 0\} \). Denote \( \sum_{i \in P} \frac{|c_i|}{2} \) by \( g \).
Then \( g = \frac{1}{2} \sum_{i \in \mathbb{P}} c_i + \frac{1}{2} \sum_{i \in \mathbb{N}} c_i \). Because \( 0 = \sum_{i \in \mathbb{N}} c_i = \sum_{i \in \mathbb{P}} c_i + \frac{1}{2} \sum_{i \in \mathbb{N}} c_i \), then \( \sum_{i \in \mathbb{P}} c_i = -\sum_{i \in \mathbb{N}} c_i \) and \( \sum_{i \in \mathbb{P}} c_i = \sum_{i \in \mathbb{N}} (-c_i) = g \).

Multiplying \( \sum_{i \in \mathbb{P}} Y_i \) by \( g \) gives \( \sum_{i \in \mathbb{P}} c_i Y_i = \frac{\sum_{i \in \mathbb{P}} \sum_{i \in \mathbb{N}} c_i (-c_i')(Y_i - Y_i')}{g} \).

However, for \( i \in \mathbb{P} \) and \( i' \in \mathbb{N} \), \( |c_i(c_i')(Y_i - Y_i')| = c_i(-c_i')|Y_i - Y_i'| \leq c_i(-c_i)c \) since by hypothesis \( |Y_i - Y_i'| \leq c \) for all \( i, i' = 1, 2, \ldots, r, i \neq i' \). Therefore,
\[
\left| \sum_{i \in \mathbb{P}} c_i Y_i \right| \leq \sum_{i \in \mathbb{P}} \frac{c \cdot g}{g} = c \cdot g = c \frac{r \cdot |c_i|}{2}.
\]

(b) If \( \sum_{1 \leq i \leq r} \frac{c_i}{2} \) for all \( c \in \mathbb{C} \), then \( |Y_i - Y_i'| \leq c \) for all \( i, i' = 1, 2, \ldots, r, i \neq i' \).

Proof: Choose \( c_i = 1, c_i' = -1, \) and \( c_j = 0, j \neq i \) or \( i' \) and the results are immediate.

Identifying \( Y_i \) with \( y_i - \mu_i, i = 1, 2, \ldots, r \) and \( c \) with \( q_{r,v}^\alpha \cdot S \), we have that \( |(y_i - y_i') - (\mu_i - \mu_i')| \leq q_{r,v}^\alpha \cdot S \) for all \( i, i' = 1, 2, \ldots, r, i \neq i' \) if and only if \( \sum_{1 \leq i \leq r} c_i (y_i - \mu_i) \leq q_{r,v}^\alpha \cdot S \cdot \frac{r \cdot |c_i|}{2} \) for all \( c \in \mathbb{C} \). This implies that \( p[|y_i - y_i' - (\mu_i - \mu_i')| \leq q_{r,v}^\alpha \cdot S] \) for all \( i, i' = 1, 2, \ldots, r, i \neq i' \) = 1 - \( \alpha \) if and only if \( p[\sum_{1 \leq i \leq r} c_i (y_i - \mu_i) \leq q_{r,v}^\alpha \cdot S \cdot \frac{r \cdot |c_i|}{2} \] for all \( c \in \mathbb{C} \) = 1 - \( \alpha \). From here, it is only a matter of a few steps of algebra to obtain statement (2).
Section 2.3  The studentized augmented range random variable and simultaneous confidence intervals for all linear combinations of means.

With contrasts the statistician is limited to making comparisons between means; he cannot make comparisons of means with theoretical values using contrasts. To make the latter types of comparisons it is necessary to move from \( \mathcal{L}_C \), the space of contrasts, to \( \mathcal{L}_L \) the space of all linear combinations of the means so that the \( \mu_i \) themselves are included in the family of statements. To make a probability statement concerning this family, the definition of the studentized augmented range random variable is needed.

Let \( y_1, y_2, \ldots, y_r \) be a random sample from a \( n(0, 1) \) distribution and \( W \) be a chi-square random variable with \( v \) degrees of freedom which is independent of \( y_i, i = 1, 2, \ldots, r \). Then \( Q_r^*, v = \max \frac{|M|_r, R_r}{\sqrt{\frac{W}{v}}} \), where \( |M|_r = \max \{|y_i|\} \), \( R_r = \max \{|y_i - y_{i'}|\} \), is the studentized augmented range random variable with \( r \) and \( v \) degrees of freedom.

As an alternative, the studentized augmented range random variable may be defined as \( Q_r^*, v = \max_{i, i'=0, \ldots, r} \frac{|y_i - y_{i'}|}{\sqrt{\frac{W}{v}}} \), where \( y_i, i = 1, 2, \ldots, r \) and \( W \) are as previously defined and \( y_0 = 0 \).
The probability statement for the family of all linear combinations of means is this:

\[(3) \ P \{ \frac{1}{r} \sum l_i \mu_i \in \frac{1}{r} \sum l_i y_i + q_{r,v} \cdot S \cdot L_\alpha \ \text{for all} \ l \in \mathcal{L} \} = 1 - \alpha, \]

where \( L_\alpha = \max \{ \sum l_i^+, -\sum l_i^- \}; \ l_i^+ = \max\{0, l_i\}, \ l_i^- = \min\{0, l_i\}, \) and \( q_{r,v} \) is the upper 100\(\alpha\) percent point of the studentized augmented range distribution with \( r \) and \( v \) degrees of freedom.

Verification of statement (3) amounts to reducing to the case of contrasts. Define \( y_0 = 0 \) and let \( \mu_0 = 0 \). The linear combination \( \sum_{i=1}^{r} l_i \mu_i \) can then be written as the contrast \( \sum_{i=0}^{r} c_i \mu_i \) where \( c_i = l_i \ i = 1, 2, \ldots, r \) and \( c_0 = -\sum l_i \). It can be shown that \( \sum_{i=0}^{r} |c_i| = 2L_\alpha \). Clearly

\[ Y_i^* = \frac{y_i - \mu_i}{\sigma}, \ i = 0, 1, \ldots, r \]

is distributed normally with mean zero and variance one. Hence, by definition,

\[ Q' = \max_{i,i'=0,\ldots,r} \frac{\| (y_i - y_{i'}) - (\mu_i - \mu_{i'}) \|}{S} \]

has a studentized augmented range distribution and

\[ P \{ \max_{i,i'=0,\ldots,r} \| (y_i - y_{i'}) - (\mu_i - \mu_{i'}) \| \leq q'_{r,v} \cdot S \} = 1 - \alpha \]

where \( q'_{r,v} \) is the upper 100\(\alpha\) percent point of the studentized augmented range distribution with \( r \) and \( v \) degrees of freedom. However, \( \max_{i,i'=0,\ldots,r} \| (y_i - y_{i'}) - (\mu_i - \mu_{i'}) \| \leq q'_{r,v} \cdot S \) if and only if \( (y_i - y_{i'}) - (\mu_i - \mu_{i'}) \leq q'_{r,v} \cdot S \).
for all \( i, i' = 0, 1, \ldots, r \). Applying lemma 1,

\[
|y_i - y_{i'}| - (\mu_i - \mu_{i'}) \leq q_{r,v}^{\alpha} \cdot S
\]

for all \( i, i' = 0, \ldots, r \) if and only if \( \sum_{i=0}^{r} c_i |y_i - \mu_i| \leq q_{r,v}^{\alpha} \cdot S \cdot \sum_{i=0}^{r} \frac{c_i}{2} \) for all \( c \in \mathcal{C} \), or equivalently, if and only if \( \sum_{i=0}^{r} c_i (y_i - \mu_i) \leq q_{r,v}^{\alpha} \cdot S \cdot \sum_{i=0}^{r} \frac{c_i}{2} \) for all \( c \in \mathcal{C} \). Therefore,

\[
p\left\{ \sum_{1}^{r} c_i (y_i - \mu_i) \leq q_{r,v}^{\alpha} \cdot S \cdot \sum_{1}^{r} \frac{c_i}{2} \right\} = 1 - \alpha.
\]

Section 2.4 Generalization of Tukey’s method to dependent random variables.

The Studentized range technique can be generalized to the case in which \((y_1, y_2, \ldots, y_r)\) has a multivariate normal distribution with mean \(\mu = (\mu_1, \mu_2, \ldots, \mu_r)\) and covariance matrix \(\sigma^2 V\) where \(V = \begin{pmatrix} 1 & \cdots & p \\ p & \ddots & \cdots \\ \vdots & \ddots & \cdots \\ p & \cdots & 1 \end{pmatrix}\).

The probability statements for the dependent case corresponding to statements (1), (2) and (3) are respectively

\[
(4) \quad p(\mu_i - \mu_{i'}, \in y_i - y_{i'}, + q_{r,v}^{\alpha} \cdot S \sqrt{1 - p}, i, i' = 1, 2, \ldots, r) = 1 - \alpha
\]

\[
(5) \quad p\left\{ \sum_{1}^{r} c_i \mu_i \in \sum_{1}^{r} c_i y_i + q_{r,v}^{\alpha} \cdot S \cdot \sqrt{1 - p} \cdot \sum_{1}^{r} \frac{c_i}{2} \right\}, \text{ for all } c \in \mathcal{C} = 1 - \alpha
\]

\[
(6) \quad p\left\{ \sum_{1}^{r} c_i \mu_i \in \sum_{1}^{r} c_i y_i + q_{r,v}^{\alpha} \cdot S \cdot \sqrt{1 - p} \cdot \sum_{1}^{r} \frac{c_i}{2} \right\}, \text{ for all } c \in \mathcal{C} = 1 - \alpha.
\]
Note that the only difference between these statements and the corresponding statements for the independent case is the insertion of $\sqrt{1 - p}$ as a factor.

The key to statements (4), (5) and (6) is in the defining of $r$ new random variables $Z_1, Z_2, \ldots, Z_r$ in terms of the $y_i$ in such a way that the $Z_i$, $i = 1, 2, \ldots, r$ are independent and, in addition, have the property that 

$$Z_i - Z_i' = y_i - y_i', \quad i, i' = 1, 2, \ldots, r$$

and

$$\eta_i - \eta_i' = \mu_i - \mu_i',$$

where $E(Z_i) = \eta_i$, $i = 1, 2, \ldots, r$. The earlier results can be applied to the $Z_i$. The details are straightforward.

Section 2.5 Applications of the studentized range technique

Tukey's studentized range technique is principally used in determining whether any of $r$ population means $\mu_1, \mu_2, \ldots, \mu_r$ differ. This is accomplished by pairwise comparisons of sample means. Quite often these means arise from one-way classification experimental design, although the method is also applicable for a two-way classification design and certain other models. Numerical illustrations for the following material are contained in Chapter V.

To see how the method works, consider a one-way classification with $r$ treatments and $n$ observations per treatment. Let $\{y_{ij}, i = 1, 2, \ldots, r, j = 1, 2, \ldots, n\}$ be $n$ independently, normally distributed random variables with common variance $\sigma^2$ and expected values $E(y_{ij}) = \mu_i$, $i = 1, 2, \ldots, r$. Let $\bar{y}_i = \frac{1}{n} \sum_{j=1}^{n} y_{ij}$ and
\[
S^2 = \sum_{ij} \frac{(y_{ij} - \bar{y}_i)^2}{r(n - 1)}.
\]
Then \(\frac{r(n - 1)S^2}{\sigma^2}\) is distributed as a chi-square with \(r(n - 1)\) degrees of freedom and \(S^2\) is independent of the sample means \(\bar{y}_i\), \(i = 1, 2, \ldots, r\). Define \(y_i^* = (\bar{y}_i - \mu_i) \sqrt{n}, i = 1, 2, \ldots, r\). Clearly \(y_i^* \sim n(0, \sigma^2)\). Therefore, by statement (1),
\[
p\{(\bar{y}_i - \mu_i) \sqrt{n} - (\bar{y}_{i'} - \mu_{i'})/\sqrt{n}\} \leq q_{\alpha}^r, r(n-1), S, i, i' = 1, \ldots, r = 1 - \alpha.
\]
Hence (8)
\[
p\{(\bar{y}_i - \mu_i) - (\bar{y}_{i'} - \mu_{i'})\} \leq q_{\alpha}^r, r(n-1), S/\sqrt{n}, i, i' = 1, 2, \ldots, r, i \neq i'.
\]
Thus we obtain a family of \((r^2)\) confidence statements with a probability error rate \(\alpha\). These statements are of the form \(\mu_i - \mu_{i'} \in \bar{y}_i - \bar{y}_{i'}, \pm q_{\alpha}^r, r(n-1), S/\sqrt{n}\) for all \(i, i' = 1, 2, \ldots, r, i \neq i'\). These confidence intervals can be used to test the null hypothesis \(H_0: \mu_i - \mu_{i'} = 0\) for all \(i, i' = 1, 2, \ldots, r, i \neq i'\). The test amounts to simply checking each of the \((r^2)\) intervals for the inclusion of zero. If zero is included in each interval accept \(H_0\); otherwise, reject \(H_0\).

In a one-way classification design the probability statement for contrasts is
\[
p\{\sum_{1}^{r} c_i \mu_i \in \sum_{1}^{r} \bar{y}_i, \pm q_{\alpha}^r, r(n-1), S/\sqrt{n}, \sum_{1}^{r} \frac{|c_i|}{2}\} = 1 - \alpha.
\]
Tukey's method is also applicable to two-way classification designs with \(r\) rows, \(c\) columns, and \(n\) observations per all. Suppose \(E(y_{ijk}) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}, i = 1, 2, \ldots, r;\)
j = 1, 2, ..., c; k = 1, 2, ..., n, and that \( \Sigma \alpha_i = \Sigma \beta_j = 0 \).

The probability statement for pairwise comparisons of row means is

\[
p\{\alpha_i - \alpha_{i'} \in \bar{y}_{i}.. \pm q_{\alpha r, rc(n-1)} S/\sqrt{cn},
\]

\[
i, i' = 1, 2, ..., r, i \neq i'
\]

where \( \bar{y}_{i}.. = \Sigma y_{ijk} \) and \( S^2 = \Sigma (y_{ijk} - \bar{y}_{ijk})^2 / rc(n - 1) \). A similar statement holds for pairwise comparison of column means.

In the special case where there is only one observation per all, there is no variation within alls, and if there is no interaction, the model becomes \( y_{ij} = \mu + \alpha_i + \beta_j \). In this case, we use \( S^2 = \Sigma (y_{ij} - \bar{y}_{i}.. - \bar{y}_{i'}.. + \bar{y}..)^2 / (r - 1)(c - 1) \) as an estimate of the error and the simultaneous confidence intervals are of the form

\[
\alpha_i - \alpha_{i'} \in \bar{y}_{i}.. - \bar{y}_{i'}.. \pm q_{r, (r-1)(c-1)} \frac{S_i}{\sqrt{c}}, i, i' = 1, 2, ..., r. \]

A similar statement holds for column comparisons.

The probability statement for column contrasts is

\[
p\{\sum_{l=1}^{c} c_j \beta_j \in \sum_{l=1}^{c} c_j \bar{y}_{..j} \pm q_{c, rc(n-1)} S_0 \sum_{l=1}^{c} \frac{|c_i|}{2} \} = 1 - \alpha \]

for all \( c \in L_c \) or if there is only one observation per all we have

\[
(10) \ p\{\sum_{l=1}^{c} c_j \beta_j \in \sum_{l=1}^{c} \bar{y}_{..j} \pm q_{c, (r-1)(c-1)} \frac{S_0}{\sqrt{c}} \sum_{l=1}^{c} \frac{|c_i|}{2} \} = 1 - \alpha.
\]
It is important to note that for Tukey's technique to be applicable it is necessary to have equal numbers of observations per all or per treatment. This requirement is one of the disadvantages of this method since often, through no fault of the experimenter, it is not possible to have equal numbers of observations per all.
CHAPTER III
Scheffé F Projections

Section 3.1 Introduction

The second technique for simultaneous inference to be considered is Scheffé F projections. This technique is widely applicable.

Let \( Y = (y_1, y_2, \ldots, y_n) \) be a vector of \( n \) independently, normally distributed random variables with common variance \( \sigma^2 \). Let the mean vector \( \mu \) be given by \( \mu = X \beta \)

where \( X = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix} \), rank \( X = p \) (\( p < n \)), and \( \beta' = (\beta_1, \beta_2, \ldots, \beta_p) \). \( X \) is a matrix of constants, but \( \beta_i, i = 1, 2, \ldots, p \) are unknown parameters and \( \sigma^2 \) is unknown.

The least squares and maximum likelihood estimators of \( \beta \) is \( \hat{\beta} = (X'X)^{-1}X'Y \). \( S^2 = \frac{Y'(I - X(X'X)^{-1}X')Y}{n-p} \)

The estimator for \( \sigma^2 \) is such that \( \frac{(n-p)S^2}{\sigma^2} \) is distributed as a chi-square with \( n - p \) degrees of freedom and \( \hat{\beta} \) and \( S^2 \) are independent. Also \( \hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1}) \).

Let \( \mathcal{L} = \{ \ell = (\ell_1, \ell_2, \ldots, \ell_p) \} \) be any fixed \( d \)-dimensional linear subspace of \( p \)-dimensional space. The Scheffé method gives simultaneous confidence intervals for the linear combinations, \( \ell'\beta = \sum_{i=1}^{p} \ell_i \beta_i \) for all \( \ell \in \mathcal{L} \). The following probability statement governs the family of all
these linear combinations:

\[ p[ \lambda' \hat{\beta} \in \mathcal{A} \setminus \hat{\lambda} + \sqrt{\frac{dF_{\alpha}}{d,n-p}} \cdot S(\hat{\lambda}'(X'X)^{-1}\hat{\lambda})^{1/2} \text{ for all } \lambda \in \mathcal{A} = 1 - \alpha \]

where \( F_{\alpha}^d,n-p \) is the upper 100\( \alpha \) percent point of the \( F \) distribution with \( d \) and \( n-p \) degrees of freedom. It should be noted that \( S^2 \hat{\lambda}'(X'X)^{-1}\hat{\lambda} \) is simply the estimate of \( \text{var}(\hat{\lambda}'\hat{\beta}) = \sigma^2 \hat{\lambda}'(X'X)^{-1}\hat{\lambda} \).

The following lemma is used in the verification of statement (1).

**Lemma 2:** For \( c > 0 \),
\[
\left| \sum_1^d a_i y_i \right| \leq c \left( \sum_1^d a_i^2 \right)^{1/2}
\]
for all \( (a_1, a_2, \ldots, a_d) \) if and only if \( \sum_1^d y_i^2 \leq c^2 \).

**Theorem:** If \( Y \sim N(X\beta, \sigma^2I) \) where \( X \) is an \( n \times p \) matrix of rank \( p (p < n) \) and \( \mathcal{A} \) is a \( d \)-dimensional subspace of \( p \)-dimensional Euclidean space, then

\[ p[ \lambda' (\hat{\beta} - \beta) \mid \lambda \in \mathcal{A} \setminus \hat{\lambda} \leq \sqrt{\frac{dF_{\alpha}}{d,n-p}} \cdot S \sqrt{\lambda'}(X'X)^{-1}\lambda, \text{ for all } \lambda \in \mathcal{A} \] = 1 - \alpha.

**Proof:** Let \( L \) be a \( d \times p \) matrix whose \( d \) linearly independent rows form a basis for \( \mathcal{A} \). Let \( \mathcal{A} = L\beta \) and \( \hat{\lambda} = L\hat{\beta} \). Then for each \( \lambda \in \mathcal{A} \), there exists \( \lambda' = (\lambda_1, \lambda_2, \ldots, \lambda_d) \) such that \( \lambda' \hat{\lambda} = \lambda' \hat{\beta} \) (\( \lambda'L = \lambda' \)) because the rows of \( L \) form a basis for \( \mathcal{A} \). Furthermore, for each \( \lambda \) there is a unique \( \lambda' \) such that \( \lambda'\beta = \lambda'\mathcal{A}' \).

\[ \hat{\mathcal{A}} = L\hat{\beta} \sim N(L\beta, \sigma^2 L(X'X)^{-1}L') = N(\mathcal{A}, \sigma^2 L(X'X)^{-1}L') \]
since $\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$. The linear combination

$$X'\beta = \lambda' \hat{\beta} \sim N(\lambda' \beta, \sigma^2 \lambda'(X'X)^{-1}\lambda) = N(\lambda' \beta, \sigma^2 \lambda' L(X'X)^{-1} L' \lambda).$$

Since $L(X'X)^{-1} L'$ is positive definite, there exists a $d \times d$ non-singular matrix $P$ such that $P(L(X'X)^{-1} L')P' = I$.

Next define $D = P \hat{\beta}$. $\hat{\beta} = P \beta \sim N(D, \sigma^2 I)$, i.e.,

$$\hat{D} = (\hat{D}_1, \hat{D}_2, \ldots, \hat{D}_p)$$

where $\hat{D}_i \sim N(D_i, \sigma^2)$. $\frac{d}{\sigma^2} \sum_i (\hat{D}_i - D_i)^2$

is distributed as a chi-square with $d$ degrees of freedom and since $\hat{\beta}$ and $S^2$ are independent, $S^2$ is independent of

$$P L \hat{\beta} = \hat{D}$$

and hence independent of $\frac{d}{\sigma^2} \sum_i (\hat{D}_i - D_i)^2$. Therefore,

$$\frac{d}{\sigma^2} \sum_i (\hat{D}_i - D_i)^2 \sim F(d, n - p).$$

We are now ready to apply lemma 2 with $c^2$ being identified with $F_{d, n-p}^\alpha \cdot d \cdot S^2$ and $\Sigma y_i^2$ being identified with $d \sum (\hat{D}_i - D_i)^2$. We obtain $\frac{d}{\sigma^2} \sum (\hat{D}_i - D_i)^2 \leq F_{d, n-p}^\alpha \cdot d \cdot S^2$ if and only if $|\sum a_i (\hat{D}_i - D_i)| \leq \sqrt{d F_{d, n-p}^\alpha} \cdot S \sqrt{\sum a_i^2}$ for all $a_1, a_2, \ldots, a_d = a'$. Therefore, $p(\frac{d}{\sigma^2} \sum (\hat{D}_i - D_i)^2 \leq F_{a, n-p}^\alpha \cdot d \cdot S^2) = 1 - \alpha$ if and only if

$$(2a) \quad p\{|a'(\hat{D} - D)| \leq \sqrt{d F_{d, n-p}^\alpha} \cdot S \sqrt{a'a} \quad \text{for all } a \in \mathcal{L}\} = 1 - \alpha.$$
\[ a'\hat{\beta} = a'P\hat{\beta} = a'PL\hat{\beta} \]
\[ a'a = a'La = a'PL(X'X)^{-1}L'P'a. \]

Because \( P \) is a non-singular matrix, there exists a one-to-one between \( a \) and \( \lambda \) by \( \lambda' = a'P \). Hence (2) can be written as
\[ P\{ |\lambda'(L\hat{\beta} - L\beta)| \leq (dF_{d,n-p}^{\alpha})^{1/2} \cdot S \cdot (\lambda'La^{-1}L'\lambda)^{1/2} \]
for all \( \lambda \) = 1 - \( \alpha \). Equivalently \( P\{ |\lambda'L(\hat{\beta} - \beta)| \leq (dF_{d,n-p}^{\alpha})^{1/2} \cdot S \cdot (\lambda'La^{-1}L'\lambda)^{1/2} \) for all \( \lambda \) = 1 - \( \alpha \) or
\[ P\{ |\lambda'(\beta - \beta)| \leq (dF_{d,n-p}^{\alpha})^{1/2} \cdot S \cdot (\lambda'(X'X)^{-1}L_L)^{1/2} \}
for all \( \lambda \) = 1 - \( \alpha \) since \( \lambda' = \lambda'L \).

Section 3.2 Extension of the Scheffé method to dependent random variables.

The Scheffé method can be extended to include random variables with an arbitrary covariance matrix \( \sigma^2V \). The method is to reduce the covariance matrix to the identity matrix by a non-singular transformation and then to apply the above theorem. Doing this gives
\[ P\{ |\lambda'\beta - \beta| \leq (dF_{d,n-p}^{\alpha})^{1/2} \cdot S \cdot (\lambda'(X'V^{-1}X)^{-1}L_L)^{1/2} \]
for all \( \lambda \) = 1 - \( \alpha \).

Section 3.3 Scheffé F projections and hypothesis testing

Thus far we have been concerned with confidence intervals; however, the Scheffé technique can also be used to test hypotheses. If linear hypotheses on \( \beta \) are to be tested, the corresponding simultaneous intervals furnish
a test which is equivalent to the likelihood ratio test of the linear hypotheses.

Consider \( d(d \leq p) \) linearly independent linear combinations \( l_i = (l_{i1}, l_{i2}, \ldots, l_{ip}), i = 1, 2, \ldots, d \) and the matrix \( L \) whose rows are these linear combinations:

\[
L = \begin{pmatrix}
  l_{11} & l_{12} & \cdots & l_{1p} \\
  \vdots & \vdots & & \vdots \\
  l_{d1} & \cdots & & l_{dp}
\end{pmatrix}
\]

Because the rows are linearly independent the rank of \( L \) is \( d \). A linear hypothesis on the regression parameters \( \beta \) has the form \( H_0: L\beta = \varphi_0 \) where \( \varphi_0 = (\varphi_{10}, \varphi_{20}, \ldots, \varphi_{d0}) \), \( \varphi_{i0} \) constant, \( i = 1, 2, \ldots, d \). \( \hat{L}\beta \sim N(L\beta, \sigma^2L(X'X)^{-1}L') \) and the likelihood ratio test of \( H_0 \) is identical to the \( F \) test based on \( \hat{L}\beta \).

When \( H_0 \) is true \( \hat{L}\beta - \varphi_0 \sim N(0, \sigma^2L(X'X)^{-1}L') \).

Furthermore, \( \frac{(\hat{L}\beta - \varphi_0)'(L(X'X)^{-1}L')(\hat{L}\beta - \varphi_0)}{d\sigma^2} \) is distributed as a chi-square with \( d \) degrees of freedom if and only if \( \frac{L(X'X)^{-1}L'}{\sigma^2} \) is idempotent of rank \( d \).

Clearly this product is equal to \( I_{d \times d} \) which is idempotent of rank \( d \). Therefore since \( S^2 \) and \( \hat{\beta} \) are independent

\[
\frac{(\hat{L}\beta - \varphi_0)'(L(X'X)^{-1}L')(\hat{L}\beta - \varphi_0)}{dS^2} \sim F(d, n - p).
\]

The test (which is identical to the likelihood ratio test) is
\( \frac{1}{d} (L\hat{\beta} - \hat{\theta})' (L(X'X)^{-1}L')^{-1} (L\hat{\beta} - \hat{\theta}) \) \begin{cases} > F_{d, n-p}^\alpha & \text{reject } H_0 \\ \leq F_{d, n-p}^\alpha & \text{accept } H_0 \end{cases}

If \( \hat{\theta} = L\hat{\beta} \), \( E_{\hat{\theta}} = \{ \hat{\theta} | (\hat{\theta} - \hat{\theta})'(L(X'X)^{-1}L')^{-1}(\hat{\theta} - \hat{\theta}) \leq dF_{d, n-p}^\alpha S^2 \} \) is the equation for an ellipsoid centered at \( \hat{\theta} \). The above test is then equivalent to the following procedure: if \( \hat{\theta} \in E_{\hat{\theta}} \) accept \( H_0 \); otherwise reject \( H_0 \). We now examine the relationship of the above test to Scheffé simultaneous intervals.

If \( \mathcal{L} \) is the linear space which the rows of \( L \) span, consider the family of simultaneous confidence intervals:
\( \ell' \beta \in \ell' \hat{\beta} \pm (dF_{d, n-p}^\alpha)^{1/2} S (\ell'(X'X)^{-1} \ell)^{1/2} \) for all \( \ell \in \mathcal{L} \).

Any \( \ell \in \mathcal{L} \) is of the form, \( \ell = \Sigma l_i l_i^1 \) where \( l_i^1 \) is the \( i \)th row of \( L \). If \( H_0 : L\beta = \theta_\circ \) is such that the hypothesized value for \( \ell_i^1 \beta \) is \( \theta_\circ_i \), then the hypothesized value for
\( \ell' \beta = \Sigma l_i \ell_i^1 \beta, \ell \in \mathcal{L}, \) is \( \theta_\circ' \ell = \Sigma l_i \theta_\circ_i \). Now \( \theta_\circ' \ell \in \ell' \hat{\beta} \pm (dF_{d, n-p}^\alpha)^{1/2} S (\ell'(X'X)^{-1} \ell)^{1/2} \) for all \( \ell \in \mathcal{L} \) if and only if \( \theta_\circ' \ell \in E_{\theta} \), (this is essentially what we proved in the theorem). Therefore the \( F \) test is equivalent to checking whether \( \theta_\circ' \ell \in \ell' \hat{\beta} \pm (dF_{d, n-p}^\alpha)^{1/2} S (\ell'(X'X)^{-1} \ell)^{1/2} \) for all \( \ell \in \mathcal{L} \). However, because it is impossible to check an infinite number of confidence intervals, the equivalent \( F \) test given above must be used.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Section 3.4 Applications

The Scheffé technique is widely applicable. In order to apply it, one simply has to specify the matrix $X$ and the linear space $L$. Since there are few restrictions on $X$, there are various possibilities for its form. This accounts for the versatility of the technique.

$X$ can be classified into one of three general types according to the nature of the variables which comprise it: qualitative, quantitative, or both qualitative and quantitative. The quantitative matrix is the one encountered in analysis of variance; its variables are simply indicators used to insert or to leave out a parameter in the mean of an observation. Its entries are 0 or 1. Quantitative matrices are usually encountered in regression problems; entries in this type of matrix are variables which are actually measurements on physical entities.

3.4.1 Application to one-way design

Let us first consider a one-way classification design with $r$ treatments with $n_i$ observations for the $i^{th}$ treatment. Let $\{y_{ij}, i = 1, 2, \ldots, r, j = 1, 2, \ldots, n_i\}$ be $r$ independent samples of independently, normally distributed random variables each with variance $\sigma^2$ and such that $E(y_{ij}) = \mu_i$, $i = 1, 2, \ldots, r$. If $r = 3$, $n_1 = 2$, $n_2 = 3$ and $n_3 = 2$, then $Y' = (y_{11}, y_{12}, y_{21}, y_{22}, y_{23}, y_{31}, y_{32})$. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\[
X' = \begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}, \quad \text{and } \beta' = (\mu_1, \mu_2, \mu_3) \text{ where}
\]

\[
\hat{\mu}_i = \bar{y}_i = \frac{1}{n_i} \sum_j y_{ij}, \quad i = 1, 2, 3. \quad \text{The } \hat{\mu}_i \text{ are independent and}
\]

\[
\text{var}(\hat{\mu}_i) = \frac{\sigma^2}{n_i}, \quad i = 1, 2, 3. \quad S^2 = \frac{3}{n} \sum \frac{n_i}{\sum (y_{ij} - \bar{y}_i)^2}{4}
\]

is distributed as a chi-square with 4 degrees of freedom and \( S^2 \) is independent of \( \hat{\mu}_1, \hat{\mu}_2, \text{ and } \hat{\mu}_3. \)

In general, \( \hat{\beta'} = (\hat{\mu}_1, \ldots, \hat{\mu}_r) \) where \( \hat{\mu}_i = \bar{y}_i = \frac{1}{n_i} \sum_j y_{ij}, \quad i = 1, 2, \ldots, r. \quad \text{The } \hat{\mu}_i \text{ are independent and}
\]

\[
\text{var}(\hat{\mu}_i) = \frac{\sigma^2}{n_i} \cdot S^2 = \sum \frac{(y_{ij} - \bar{y}_i)^2}{N - r}
\]

is distributed as a chi-square with \( N - r \) degrees of freedom and is independent of the \( \hat{\mu}_i. \)

In a one-way classification, pairwise mean comparisons and contrasts are most often of concern. Since the linear space \( \mathcal{X}_c \) is spanned by \( \mathcal{C}_1 = (1, -1, \ldots, 0), \)
\( \mathcal{C}_2 = (0, 1, -1, \ldots, 0), \ldots, \mathcal{C}_{r-1} = (0, 0, \ldots, 1, -1), \) the dimension of \( \mathcal{X}_c \) is \( r - 1. \)

\[
\text{var} \left( \sum_{C_i} \hat{\mu}_i \right) = \sum_{C_i} \text{var}(C_i \hat{\mu}_i) = \sum_{C_i} \frac{\sigma^2}{n_i} = \sigma^2 \sum_{C_i} \frac{c_i^2}{n_i}
\]

and \( \text{var} \left( \sum_{C_i} \mu_i \right) = S^2 \sum_{C_i} \frac{c_i^2}{n_i}. \) Therefore from statement (1),

we have
for all \( c \in \mathcal{C} \). Observe that unlike Tukey's technique, the Scheffé method does not require equal numbers of observations per treatment.

Section 3.4.2 Application to two-way classification

Next let us consider how the Scheffé technique can be applied to a two-way classification with interactions present, \( r \) rows, \( c \) columns, and \( m \) observations per cell.

Let \( Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \), \( i = 1, 2, \ldots, r \), \( j = 1, 2, \ldots, c \), and \( k = 1, 2, \ldots, m \), where \( 0 = \Sigma \alpha_i = \Sigma \beta_j = \Sigma (\alpha\beta)_{ij} = \Sigma (\alpha\beta)_{ij} \) and \( \epsilon_{ijk} \sim n(0, \sigma^2) \). Also, assume that the \( \epsilon_{ijk} \) are independent.

The least squares estimators for the parameters are

\[
\hat{\mu} = \frac{\Sigma Y_{ijk}}{rcm} = \overline{Y}...; \quad \hat{\alpha}_i = \overline{Y}_i... - \overline{Y}... \quad \text{where} \quad \overline{Y}_i... = \frac{\Sigma Y_{ijk}}{cm},
\]

\( i = 1, 2, \ldots, r \); \( \hat{\beta}_j = \overline{Y}_j... - \overline{Y}... \quad \text{where} \quad \overline{Y}_j... = \frac{\Sigma Y_{ijk}}{rm}, \)

\( j = 1, 2, \ldots, c \); \( (\hat{\alpha}\hat{\beta})_{ij} = \overline{Y}_{ij}... - \overline{Y}_i... - \overline{Y}_j... + \overline{Y}... \quad \text{where} \quad \Sigma_{k=1}^{m} \frac{Y_{ijk}}{m} = \overline{Y}_{ij}... \); and \( S^2 = \frac{1}{rc(m-1)} \Sigma_{ijk} (Y_{ijk} - \overline{Y}_{ij}...)^2 \) is distributed as a chi-square with \( rc(m-1) \) degrees of freedom independently of the estimators for the parameters.

Suppose the experimenter is interested in contrasts of the row effects. For the same reason as for one-way
classification, the dimension of $\mathcal{L}_c$, the space of linear
contrasts for rows, is $r - 1$. Also, $\text{var}(\Sigma c_i \hat{\alpha}_i) = \Sigma \frac{2 \alpha^2}{cm}$; so, $\text{var}(\Sigma c_i \hat{\alpha}_i) = S \Sigma \frac{2 \alpha^2}{cm}$. Therefore, applying
statement (1) we have

$$p_c \left\{ \frac{r c_i \alpha_i}{E c_i} \in \frac{r c_i \alpha_i}{E c_i} + \left[ \left( r - 1 \right)^{F_{r-1, r^2-1}} \right]^{1/2} \cdot S \left( \frac{cm r c_i^2}{1} \right)^{1/2} \right\} = 1 - \alpha.$$ 

The probability statement for row contrasts is

$$p_c \left\{ \frac{c c_j \beta_j}{E c_j} \in \frac{c c_j \beta_j}{E c_j} + \left[ \left( c - 1 \right)^{F_{c-1, r^2-1}} \right]^{1/2} \cdot S \left( \frac{cm c_j^2}{1} \right)^{1/2} \right\} = 1 - \alpha.$$ 

In the special case where $m = 1$ and interaction is
assumed to be zero, we use $S_1^2 = \Sigma \frac{(y_{i,j} - \bar{y}_{i} - \bar{y}_{j} + \bar{y}_{..})^2}{(r - 1)(c - 1)}$ as an estimate of the error. Then the statement for row
contrasts becomes

$$p_c \left\{ \frac{r c_i \alpha_i}{E c_i} \in \frac{r c_i \alpha_i}{E c_i} + \left[ \left( r - 1 \right)^{F_{r-1, (r-1)(c-1)}} \right]^{1/2} \cdot S_1 \sqrt{\frac{r c_i^2}{1}} \right\} = 1 - \alpha$$ 

for all $c \in \mathcal{L}_c$ and similarly for column contrasts we have

$$p_c \left\{ \frac{c c_j \beta_j}{E c_j} \in \frac{c c_j \beta_j}{E c_j} + \left[ \left( c - 1 \right)^{F_{c-1, (r-1)(c-1)}} \right]^{1/2} \cdot S_1 \sqrt{\frac{r c_j^2}{1}} \right\} = 1 - \alpha.$$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Section 3.4.3 Application to regression: Simple linear regression and general regression

As was mentioned before, the Scheffé technique is also applicable to regression problems. Consider the simple linear regression model: \( y_i = \alpha + \beta x_i + e_i \), \( i = 1, 2, \ldots, n \), \( e_i \sim N(0, \sigma^2) \), and the \( e_i \) are independent.

\[ X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \]
where \( x_1, x_2, \ldots, x_n \) are constants,

\[ X'X = \begin{bmatrix} n & \Sigma x_1 \\ \Sigma x_1 & \Sigma x_1^2 \end{bmatrix} \]
and \( (X'X)^{-1} = \begin{bmatrix} \frac{1}{n} + \frac{x^2}{\Sigma (x_i - \bar{x})^2} & -\frac{x}{\Sigma (x_i - \bar{x})^2} \\ -\frac{x}{\Sigma (x_i - \bar{x})^2} & \frac{1}{\Sigma (x_i - \bar{x})^2} \end{bmatrix} \]

\( \hat{\theta} = \bar{y} - \hat{\beta} \bar{x} \) and \( \hat{\theta} = \frac{n}{\Sigma (x_i - \bar{x})^2} \sum_{i=1}^{n} \frac{(y_i - \bar{y})(x_i - \bar{x})}{x^2} \cdot (\hat{\alpha}, \hat{\beta})' \sim N((\alpha, \beta)', \Sigma) \),

\[ \sigma^2 (x'x)^{-1} \] and \( s^2 = \frac{1}{n - 2} \left[ \sum_{i=1}^{n} (y_i^2 - n\bar{y}^2 - \hat{\beta} \sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})) \right] \]
is such that \( \frac{(n - 2)s^2}{\sigma^2} \) is distributed as a chi-square with \( n - 2 \) degrees of freedom independently of \( \hat{\theta} \) and \( \hat{\beta} \).

Suppose that simultaneous confidence intervals for arbitrary linear combinations of \( \alpha \) and \( \beta \) are of concern. Then the dimension of \( \mathbf{z} \) is 2. Let \( (z_1, z_2)' = \mathbf{z} \). To make
a probability statement we need \( A'(X'X)^{-1}A \). Let 

\[
A = \frac{n}{\Sigma} (x_i - \bar{x})^2.
\]

Then \( A'(X'X)^{-1}A = (l_1, l_2) \begin{bmatrix} 1/n + \frac{\bar{x}}{A} & \frac{\bar{x}}{A} \\ -\frac{\bar{x}}{A} & 1/A \end{bmatrix} \begin{bmatrix} l_1 \\ l_2 \end{bmatrix} = l_1^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{A} \right)
\]

\[-2l_1 l_2 \frac{\bar{x}}{A} + \frac{l_2^2}{A} = \frac{l_1^2}{n \Sigma (x_i - \bar{x})^2} + \frac{(l_2 - l_1 \bar{x})^2}{\Sigma (x_i - \bar{x})^2}.\]

Therefore,

\[
(9) \quad p\{l_1 \alpha + l_2 \beta \in l_1 \hat{\alpha} + l_2 \hat{\beta} + (2F_{2, n-2})^{1/2} S \sqrt{\frac{l_1^2}{n} + \frac{(l_2 - l_1 \bar{x})^2}{\Sigma (x_i - \bar{x})^2}}
\]

for all \( \alpha \in \mathcal{E} \} = 1 - \alpha.\) The main intervals of interest in this set are those for \( \alpha, \beta, \) and \( \alpha + \beta \chi.\)

Now consider the general regression model: \( Y = x_\beta + \varepsilon, \)

\( \varepsilon \sim N(0, \sigma^2 I), \) and \( x_{nxp} \) has rank \( p(p < n). \) For a given \( \beta, \)

the function \( f(x_1, x_2, \ldots, x_p) = \beta_1 x_1 + \ldots + \beta_p x_p \) defines

a surface over the p-dimensional space of \( x_1, x_2, \ldots, x_p. \)

A surface so defined is called a regression surface.

It is often desirable to have confidence bands for the regression surfaces. This requires two functions \( f_1 \) and \( f_2 \) such that \( p\{f_1(x_1, x_2, \ldots, x_p) \leq f(x_1, x_2, \ldots, x_p) \leq f_2(x_1, x_2, \ldots, x_p) \} \) for all \( (x_1, x_2, \ldots, x_p) \) = 1 - \( \alpha. \) It is important to note that for each \( (x_1, x_2, \ldots, x_p) \) in p-dimensional space, both upper and lower confidence limits on the value of the true regression function are required.

The Scheffé technique can be used to band the entire regression surface. Of course, in most practical cases,
only part of the regression surface is of concern. In this respect, the Scheffé method is "wasteful."

A 100(1-\alpha) percent confidence band for \( \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \) is

\[
\sum_{i=1}^{p} \beta_i x_i \in \sum_{i=1}^{p} \hat{\beta}_i x_i + (pF_{p,n-p}^\alpha)^{1/2} S(x'(x'x)^{-1}x)^{1/2}
\]

for all \( x = (x_1, x_2, \ldots, x_p) \).

The proof for this band has already been derived. The linear space of concern is simply the space of all linear combinations of the coordinates of \( \beta \) which obviously has dimension \( p \).

Section 5.4.4 Application to prediction in regression analysis

Another problem that often arises in regression analysis is that of using a random sample \( y_1, y_2, \ldots, y_n \) to predict to future observations of the dependent random variables at \( k \) different values of the independent random variables. Let \( y_1^0, y_2^0, \ldots, y_k^0 \) be the "future" random variables at the points \( x_i^0 = (x_{i1}, x_{i2}, \ldots, x_{ip}) \), \( i = 1, 2, \ldots, k \). The estimators for \( y_i^0 \) are \( \hat{y}_i^0 = x_i^0 \hat{\beta}, i = 1, 2, \ldots, k \). The probability statement is

\[
\sum_{i=1}^{p} \beta_i x_i \sim n(x_i^0 \beta, \sigma^2 x_i^0 (x'x)^{-1}x_i^0), i = 1, 2, \ldots, k.
\]

The Scheffé technique can be used to bracket these future observations in simultaneous confidence intervals.

The probability statement is
(11) \[ p\{y_i^0 \in \hat{y}_i^0 \pm (kF_{k,n-p})^{1/2} \cdot s \cdot (1 + x_i^0(x'x)^{-1}x_i^0)^{1/2}, \]
i = 1, 2, \ldots, k \} = 1 - \alpha, \text{ where } y_i^0 \text{ is the predicted value.}

This probability statement gives confidence intervals for only the specified \( k \) future observations, but one could obtain simultaneous confidence intervals for all linear combinations of the \( y_i^0 \), \( i = 1, 2, \ldots, k \) with the same probability, \( 1 - \alpha \). Therefore, the probability statement (11) is strictly greater than \( 1 - \alpha \) since not all possible statements are used. The verification of statement (11) depends on the extension of the Scheffé method to dependent variables.

In the special case of simple linear regression, \( y_i = \alpha + \beta x_i \), \( i = 1, 2, \ldots, k \), statement (11) simplifies to

(12) \[ y_i^0 \in \hat{\alpha} + x_i^0\hat{\beta} \pm (kF_{k,n-2})^{1/2} \cdot s \cdot (1 + \frac{1}{n} + \frac{(x_i^0 - \bar{x})^2}{\sum (x_i - \bar{x})^2})^{1/2}, \]
i = 1, 2, \ldots, k.

That \( x_i^0(x'x)^{-1}x_i^0 = \frac{1}{n} + \frac{(x_i^0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} \), \( i = 1, 2, \ldots, k \) can be obtained immediately from the derivation of \( \hat{\beta}'(x'x)\hat{\beta} \) for statement (9). In statement (12) we are concerned with the special case when \( \hat{\beta}' = (1, x_i^0) = x_i^0' \).
CHAPTER IV

Bonferroni T Statistics and Studentized Maximum Modulus Techniques

Section 4.1 Bonferroni T statistics

Both the studentized range technique and Scheffé F projections require the random variables \( y_1, y_2, \ldots, y_r \) to be independent. The Bonferroni T statistics method does not have this requirement and is therefore a useful tool when the above methods cannot be applied. Moreover, in certain instances, it is a strong competitor with the Scheffé and Tukey techniques even when independence is present.

Suppose that \( y_1, y_2, \ldots, y_k \) are random variables such that \( y_i \sim n(\mu, \sigma_i^2), \ i = 1, 2, \ldots, k \) and that \( S_i^2 = \frac{\overline{y}_i - y_i}{\sigma_i^2} \) is such that \( \frac{\overline{y}_i - y_i}{\sigma_i^2} \) is distributed as a chi-square with \( v_i \) degrees of freedom independently of \( y_i, i = 1, 2, \ldots, k \). No assumption is made regarding independence of the \( y_i \) nor if the \( S_i^2, i = 1, 2, \ldots, k \). Clearly \( \frac{y_i - \mu_i}{\sigma_i^2} \sim n(0,1) \) and \( S_i^2 \) is independent of this ratio. Therefore,

\[
T_i = \frac{y_i - \mu_i}{S_i^2} \sim t(v_i), \ i = 1, 2, \ldots, k
\]

and

\[
p\{ |y_i - \mu_i| \leq t^{\sigma/2k} S_i \} = 1 - \alpha/k, \text{ where } t^{\sigma/2k}_v \text{ is the } 100 \, \alpha/2k \text{ percentile point of the } t \text{ distribution with } v_i
\]
degrees of freedom for \( i = 1, 2, \ldots, k \), i.e.,

\[
p\{\mu_i \in y_i \pm \frac{t_{\sigma/2k} \cdot S_i}{v_i}\} = 1 - \alpha/k, \quad i = 1, 2, \ldots, k.
\]

These, of course, are individual confidence statements and what we want is a set of simultaneous confidence intervals. To obtain them, one simply uses the Bonferroni inequality derived in the first chapter. If \( A_i \) is the event that \( \mu_i \in y_i \pm \frac{t_{\sigma/2k} \cdot S_i}{v_i} \), then \( p(A_i) = \frac{\alpha}{k}, \quad i = 1, 2, \ldots, k \).

Therefore, \( p(\cap A_i) \geq 1 - \frac{k}{i=1} \alpha/k = 1 - \alpha \), or equivalently,

\[
(1) \quad p\{\mu_i \in y_i \pm \frac{t_{\sigma/2k} \cdot S_i}{v_i}, \quad i = 1, 2, \ldots, k\} \geq 1 - \alpha.
\]

Here we have used equal significance levels for each of the individual statements. However, should greater sensitivity be desired for some intervals than for others, unequal significance levels can be used. Furthermore, any combination \( \alpha_1, \alpha_2, \ldots, \alpha_k \) where \( \sum_{i=1}^k \alpha_i = \alpha \) will produce \( \alpha \) as the bound for the probability error rate.

Consider a one-way classification with \( r \) treatments and unequal numbers of observations per treatment. Let \( \{y_{ij}, \quad i = 1, 2, \ldots, r, \quad j = 1, 2, \ldots, n_i\} \) be \( r \) samples of normally distributed random variables such that

\[
E(y_{ij}) = \mu_i \quad \text{and} \quad \text{var}(y_{ij}) = \sigma^2, \quad i = 1, 2, \ldots, r, \quad j = 1, 2, \ldots, n_i.
\]

Then \( E(\bar{y}_{i.}) = E(\Sigma j \frac{y_{ij}}{n_i}) = \mu_{ij} \quad \text{var}(\bar{y}_{i.}) = \frac{\sigma^2}{n_i} = \sigma^2_i \); and \( S^2 = \Sigma \Sigma (y_{ij} - \bar{y}_{i.})^2 \)

\[\frac{N - r}{N} \]

where \( N = \Sigma_{i=1}^r n_i \).
is such that \( \frac{(N - r)S^2}{\sigma^2} \) is distributed as a chi-square with 
\( N - r \) degrees of freedom independently of the \( y_{ij} \). Then 
letting \( S_i^2 = \frac{S^2}{n_i} \) and applying statement (1) yields the 
following probability statement:

(2) \( p \{ \mu_i \in \bar{y}_i \pm t_{N-r}^{\alpha/2k} \frac{S}{\sqrt{n_i}}, i = 1, 2, \ldots, r \} \geq 1 - \alpha \)

where \( t_{N-r}^{\alpha/2k} \) is the 100\( \frac{\alpha}{2k} \) percent point of the \( t \) distribution with \( N - r \) degrees of freedom.

Suppose that simultaneous confidence intervals are 
needed for \( k \) contrasts, \( \sum_{i=1}^{r} C_{h_i} \mu_i, h = 1, 2, \ldots, k, \) 
of the above treatment means \( E(\sum_{i=1}^{r} C_{h_i} \bar{y}_i) = \sum_{i=1}^{r} C_{h_i} \mu_i \) 
and \( \sigma_n^2 = \operatorname{var}(\sum_{i=1}^{r} C_{h_i} \bar{y}_i) = \sum_{i=1}^{r} (\operatorname{var}(C_{h_i} \bar{y}_i)) = \sigma^2 \sum_{i=1}^{r} \frac{C_{h_i}^2}{n_i}, \) 
h = 1, 2, \ldots, k. Thus the probability statement for these 
contrasts is

(3) \( p \{ \sum_{i=1}^{r} C_{h_i} \mu_i \in \sum_{i=1}^{r} C_{h_i} \bar{y}_i \pm t_{\alpha/2k} \frac{S}{\sqrt{n_i}} \sqrt{\sum_{i=1}^{r} C_{h_i}^2}, \) 
h = 1, 2, \ldots, k \} \geq 1 - \alpha \)

or in the event that \( n_i = n, i = 1, 2, \ldots, r \), this state-
ment simplifies to

(3a) \( p \{ \sum_{i=1}^{r} C_{h_i} \mu_i \in \sum_{i=1}^{r} C_{h_i} \bar{y}_i \pm t_{\alpha/2k}^{r(n-1)} \frac{S}{\sqrt{n}} \sqrt{\sum_{i=1}^{r} C_{h_i}^2}, \) 
h = 1, 2, \ldots, k \} \geq 1 - \alpha \)

where \( t_{\alpha/2k}^{r(n-1)} \) and \( t_{\alpha/2k}^{N-r} \) are the 100\( \frac{\alpha}{2k} \) percent points of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
the t distributions with \( r(n - 1) \) and \( N - r \) degrees of freedom respectively.

Bonferroni t-statistics are also applicable to an \( r \times c \) two-way classification with \( n \) observations per cell. The model is

\[
y_{ijl} = \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij} + e_{ijl}, \quad i = 1, 2, \ldots, r; \quad j = 1, 2, \ldots, c, \quad l = 1, 2, \ldots, n,
\]

where \( \Sigma_{i=1}^{r} \alpha_i = \Sigma_{j=1}^{c} \beta_j = \Sigma_{i=1}^{r} (\alpha \beta)_{ij} = 0 \) and the \( e_{ijl} \) are distributed normally and independently with mean zero and variance \( \sigma^2 \). If \( k \) row contrasts are of concern then the appropriate probability statement is

\[
(4) \quad \Pr \left\{ \frac{r}{\sum_{i=1}^{r} C_{hi} \alpha_i} \leq \frac{r}{\sum_{i=1}^{r} C_{hi} \bar{y}_i} - t_{rc(n-1)}^{r/2k} \left( \frac{1}{cn} \sum_{l=1}^{r} \frac{C_{hl}^2}{2} \right)^{1/2}, \quad h = 1, 2, \ldots, k \right\} \geq 1 - \alpha
\]

where \( t_{rc(n-1)}^{r/2k} \) is the \( 100\alpha/2k \) percent point of the t distribution with \( rc(n - 1) \) degrees of freedom. A similar statement holds for column contrasts.

In the event that there is only one observation per cell and assuming that the interaction is zero, the model becomes

\[
y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \quad i = 1, 2, \ldots, r, \quad j = 1, 2, \ldots, c\]

where \( \Sigma_{i=1}^{r} \alpha_i = \Sigma_{j=1}^{c} \beta_j = 0 \) and we use

\[
S^2 = \frac{\sum_{ij} \left( y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y} \right)^2}{(r - 1)(c - 1)}
\]

as an estimate of the error. The \( 1 - \alpha \) percent simultaneous confidence intervals for \( k \) column contrasts in this case are

\[
\left( \sum_{j=1}^{c} C_{hj} \beta_j \leq \sum_{j=1}^{c} C_{hj} \bar{y}_j \right) \pm
\]
\[ t^{(r-l)(c-l)}_{(r-l)(c-l)} \left( \frac{1}{r} \sum_{j=1}^{c} c_h j^2 \right)^{1/2}, \ h = 1, 2, \ldots, k, \] where \( t^{(r-l)(c-l)}_{(r-l)(c-l)} \) is the \( 100\frac{\alpha}{2k} \) percent point of the t distribution with \( (r - 1)(c - 1) \) degrees of freedom.

It is important to note here that we have obtained simultaneous confidence intervals for only a finite number of contrasts; whereas, the Tukey and Scheffé method give simultaneous confidence intervals for all possible contrasts. Observe above that as \( k \) increases so does \( t_{d.f.}^{a/2k} \) in each case, giving wider confidence intervals. Thus, one would expect shorter confidence intervals when using Bonferroni t-statistics if only a few contrasts are of concern than he would expect for the same contrasts using the Scheffé or Tukey techniques.

Section 4.2 Studentized Maximum Modulus Technique

Another technique for simultaneous inference is the studentized maximum modulus technique. One of its drawbacks is that it has not been extensively tabulated.

The studentized maximum modulus statistic with \( K, v \) degrees of freedom is

\[ \frac{|M|_k}{\sqrt{\frac{\chi^2_v}{v}}} \]

absolute value of \( k \) independent random variables which are normal with mean zero and variance one, and \( \chi^2_v \) is an independent chi-square with \( v \) degrees of freedom.

Let \( y_i \sim n(\mu_i, d_i \sigma^2) \), \( i = 1, 2, 3, \ldots, K \) be \( k \)
independent random variables, where the \( d_i \) are known constants but \( \mu_i \) and \( \sigma^2 \) are unknown for \( i = 1, 2, \ldots, K \). Also, let \( S^2 \) be a chi-square estimator of \( \sigma^2 \) with \( v \) degrees of freedom which is independent of the \( y_i \). The probability statement for the maximum modulus technique is

\[
p\{\mu_i \in y_i \pm |M|^\alpha K_v \sqrt{d_i} S, i = 1, 2, \ldots, K\} = 1 - \alpha.
\]

This method can also be extended to arbitrary linear combinations.

Independence of the numerators is required, but unlike Tukey's method, equal numbers of observations per cell is not required.

Section 4.3 Multiple Range Tests

The name "multiple range tests" applies to a number of techniques of simultaneous inference. They are mentioned briefly here because of their historical significance: They are among the first tools of simultaneous statistical inference to be developed. As the name implies these multiple range tests can be used only in testing hypotheses; they cannot be used for obtaining simultaneous confidence intervals.

Multiple range tests are used primarily in testing mean differences in balanced one-way, two-way, etc., classifications. "The basic idea in these tests is the same: as some means are declared significant, the critical point for significance of the remaining means is decreased
to conform with the size of the remaining group."
CHAPTER V
Numerical Examples

Section 5.1 Application of the Scheffé, Tukey, and Bonferroni techniques to a randomized complete block design.

It is the purpose of this last chapter to illustrate numerically some of the techniques derived in the preceding chapters. We first consider a randomized complete block design. The following table gives the results of an experiment designed to compare the yields of five different kinds of wheat measured in bushels.

<table>
<thead>
<tr>
<th>VARIETIES</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks 1</td>
<td>61</td>
<td>60</td>
<td>64</td>
<td>69</td>
<td>83</td>
</tr>
<tr>
<td>Blocks 2</td>
<td>65</td>
<td>66</td>
<td>66</td>
<td>80</td>
<td>83</td>
</tr>
<tr>
<td>Blocks 3</td>
<td>60</td>
<td>55</td>
<td>68</td>
<td>72</td>
<td>70</td>
</tr>
<tr>
<td>Blocks 4</td>
<td>75</td>
<td>70</td>
<td>80</td>
<td>80</td>
<td>89</td>
</tr>
</tbody>
</table>

Means: \( \overline{y}_1 = 65.3 \) \( \overline{y}_2 = 62.7 \) \( \overline{y}_3 = 69.5 \)
\( \overline{y}_4 = 75.3 \) \( \overline{y}_5 = 81.3 \) \( \overline{y}_{..} = 70.8 \)

ANOVA

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>3</td>
<td>553.2</td>
<td>184.4</td>
<td></td>
</tr>
<tr>
<td>Varieties</td>
<td>4</td>
<td>905.2</td>
<td>226.3</td>
<td>16.88</td>
</tr>
<tr>
<td>Residual</td>
<td>12</td>
<td>160.8</td>
<td>13.4</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>1619.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The appropriate model then is

\[
y_{ij} = \mu + \beta_i + \gamma_j + \epsilon_{ij},
\]

\(i = 1, 2, 3, 4, j = 1, 2, \ldots, 5\), where \(\Sigma \beta_i = \Sigma \gamma_j = 0\).

\[
\hat{B}_i = \bar{y}_i - \bar{y}, \quad \hat{\gamma}_j = \bar{y}_j - \bar{y}, \quad \text{and} \quad S_I^2 = \frac{\sum_{i=1}^{4} \sum_{j=1}^{5} (y_{ij} - \bar{y}_i + \bar{y}_j - \bar{y})^2}{4.3}.
\]

Since 16.88 > \(F_{4,12}^{0.05} = 3.26\), we would reject the hypothesis that all varieties are equal and consider then differences in varieties.

Suppose that 95% simultaneous confidence intervals are desired for pairwise differences in variety effects. The probability statement for contrasts using the Sheffé technique is

\[
P\left[\sum_{c=1}^{\mathcal{L}_c} c J_c \epsilon \sum_{c=1}^{\mathcal{L}_c} \gamma_c^* \bar{y}_c \right] + [(c - 1)F_{c-1}^\alpha (r-1)(c-1)]^{1/2} \leq \frac{S_I}{\sum_{r} c^2} \left(\sum_{c=1}^{\mathcal{L}_c} c \right)^{1/2} \text{ for all } c \in \mathcal{L}_c = 1 - \alpha.
\]

\[
\hat{\gamma}_1 = 65.3 - 70.8 = -5.5, \quad \hat{\gamma}_4 = 75.3 - 70.8 = 4.5
\]

\[
\hat{\gamma}_2 = 62.7 - 70.8 = -8.1, \quad \hat{\gamma}_5 = 81.3 - 70.8 = 10.5
\]

\[
\hat{\gamma}_3 = 69.5 - 70.8 = -1.3
\]

Here \(\sigma = 4\), \(c = 5\), \(S_I^2 = 13.4\), and \(S_I = 3.66\).

\[
F_{4,12}^{0.05} = 3.26, \quad \sqrt{\frac{\mathcal{L}_c}{4}} = \sqrt{13.04} = 3.61.
\]

We first look at pairwise differences in varieties, \(J_j - J_j'\). For these pairwise differences in varieties

\[
\sqrt{\frac{4}{\sum_{j=1}^{5} c_j^2}} = \sqrt{2/4} = .71. \quad \text{Therefore}, \quad \left[4F_{4,12}^{0.05}\right]^{1/2} S_I \sqrt{\frac{4}{\sum_{j=1}^{5} c_j^2}} = 9.38 \text{ for these contrasts. The simultaneous confidence intervals are as follows:}
\]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Variety difference

<table>
<thead>
<tr>
<th>Difference</th>
<th>Estimate</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J_1 - J_2$</td>
<td>$\hat{J}<em>j - \hat{J}</em>{j'}$</td>
<td>$\pm 9.38$</td>
</tr>
<tr>
<td>$J_1 - J_3$</td>
<td>$\pm 9.58$</td>
<td></td>
</tr>
<tr>
<td>$J_1 - J_4$</td>
<td>$\pm 9.58$</td>
<td></td>
</tr>
<tr>
<td>$J_1 - J_5$</td>
<td>$\pm 9.58$</td>
<td></td>
</tr>
<tr>
<td>$J_2 - J_3$</td>
<td>$\pm 9.58$</td>
<td></td>
</tr>
<tr>
<td>$J_2 - J_4$</td>
<td>$\pm 9.58$</td>
<td></td>
</tr>
<tr>
<td>$J_2 - J_5$</td>
<td>$\pm 9.58$</td>
<td></td>
</tr>
<tr>
<td>$J_3 - J_4$</td>
<td>$\pm 9.58$</td>
<td></td>
</tr>
<tr>
<td>$J_3 - J_5$</td>
<td>$\pm 9.58$</td>
<td></td>
</tr>
<tr>
<td>$J_4 - J_5$</td>
<td>$\pm 9.58$</td>
<td></td>
</tr>
</tbody>
</table>

Suppose now that the hypothesis $H_0: J_j - J_{j'} = 0$, $j \neq j'$, $j, j' = 1, 2, \ldots, 5$, is to be tested. Recalls that testing this hypothesis is equivalent to checking each of the above confidence intervals for the inclusion of zero. The intervals which do not include zero have a * beside them. Hence, in this case, we would reject $H_0$ and conclude that $J_4 > J_1$, $J_5 > J_1$, $J_4 > J_2$, $J_5 > J_2$ and $J_5 > J_3$.

In addition, we wish to compute a confidence interval for $J_1 + J_2 - (J_3 + J_4)$, simply compute $\frac{\sum c_i^2}{5} = \frac{4}{4} = 1$. Then $\sqrt{\sum c_i^2} = 1$ and a 95% confidence interval for $J_1 + J_2 - (J_3 + J_4)$ is therefore, $-5.5 - 8.1 + 1.3 - 4.5 + 13.01$; i.e., $(-29.81, -3.79)$. To test $H_0: J_1 + J_2 - (J_3 + J_4) = 0$, simply observe that zero is not in the...
Next consider 95 o/o simultaneous confidence intervals for pairwise differences in variety effects using Tukey's method. The probability statement in this case is
\[ \Pr \left\{ \sum_{j=1}^{c} c_j \hat{J}_j - \sum_{j=1}^{c} c_j \hat{J}_j \geq q_c, (r-1)(c-1) \frac{S_I}{\sqrt{r}} \frac{c}{\sqrt{2}} \right\} = 1 - \alpha. \]
Here \( c = 5, r = 4, q_{5,12}^{0.05} = 4.508, S_I = 3.66. \) and \( \frac{S_I}{\sqrt{4}} = 1.38. \) Therefore, \( q_{5,12}^{0.05} \frac{S_I}{\sqrt{4}} = 8.25 \) is the critical constant for pairwise differences in varieties since \( \frac{c}{\sqrt{2}} \frac{1}{\sqrt{2}} = 1 \) in this case. The desired confidence intervals are:

<table>
<thead>
<tr>
<th>Variety difference</th>
<th>( \hat{J}_j - \hat{J}_j \pm 8.25 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( J_1 - J_2 )</td>
<td>((-5.65, 10.85))</td>
</tr>
<tr>
<td>( J_1 - J_3 )</td>
<td>((-10.45, 4.05))</td>
</tr>
<tr>
<td>( J_1 - J_4 )</td>
<td>((-18.25, -1.75))*</td>
</tr>
<tr>
<td>( J_1 - J_5 )</td>
<td>((-24.25, -7.75))*</td>
</tr>
<tr>
<td>( J_2 - J_3 )</td>
<td>((-15.05, 1.45))</td>
</tr>
<tr>
<td>( J_2 - J_4 )</td>
<td>((-20.85, -3.25))*</td>
</tr>
<tr>
<td>( J_2 - J_5 )</td>
<td>((-26.85, -9.35))*</td>
</tr>
<tr>
<td>( J_3 - J_4 )</td>
<td>((-14.05, 2.45))</td>
</tr>
<tr>
<td>( J_3 - J_5 )</td>
<td>((-20.05, -3.55))*</td>
</tr>
<tr>
<td>( J_4 - J_5 )</td>
<td>((-14.25, 2.25))</td>
</tr>
</tbody>
</table>

The intervals with * beside them do not include zero and as before we would reject \( H_0 \) and conclude that \( J_4 > J_1, J_5 > J_1, J_4 > J_2, J_5 > J_2, \) and \( J_5 > J_3. \)
If we wish to compute an interval for the contrast
\[ J_1 + J_2 = (J_3 + J_4), \]
simply compute \( \frac{5}{2} \frac{|c_i|}{2} = 2. \) Then
\[ q_{.05}^{12} \frac{S}{\sqrt{4}} = 16.5 \] and we obtain the following confidence interval:
\[ 65.3 + 62.7 - 69.5 - 75.3 + 16.5 = (33.3, -3.3). \]

To test \( H_0: J_1 + J_2 - (J_3 + J_4) = 0, \) simply observe that zero is not in the confidence interval for the contrast; therefore, reject \( H_0. \)

Finally, consider the Bonferroni method for obtaining a set of 95% confidence intervals for pairwise differences in variety means. The probability statement we need is
\[ p\left( \sum c_{ij} \overline{y}_i - \sum c_{ij} \overline{y}_j \right) \leq t^{\alpha/2k} (r-1)(c-1) \frac{S_i}{\sqrt{\sum c_{ih_i}^2}}^{1/2}, \]
for \( h = 1, 2, \ldots, K \geq 1 - \alpha \)

In this example, \( \alpha = .05, k = 10, r = 4, \) and \( c = 5. \)
\[ t_{.05/2(10)} = 3.43, S_i = 3.66 \]
and for pairwise differences of means \( \left( \frac{1}{r} \sum c_{hi} \right)^{1/2} = (\frac{1}{4})^{1/2} = .71. \) Thus, the critical value is \( (3.43)(3.66)(.71) = 8.91. \)

The simultaneous confidence intervals are as follows:

<table>
<thead>
<tr>
<th>Variety difference</th>
<th>( \hat{J}_j - \hat{J}_j \pm 8.91 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( J_1 - J_2 )</td>
<td>(-6.31, 11.51)</td>
</tr>
<tr>
<td>( J_1 - J_3 )</td>
<td>(-13.11, 4.71)</td>
</tr>
<tr>
<td>( J_1 - J_4 )</td>
<td>(-18.91, -1.09)</td>
</tr>
<tr>
<td>( J_1 - J_5 )</td>
<td>(-24.91, -7.09)</td>
</tr>
</tbody>
</table>
\[
\begin{array}{ll}
J_2 - J_3 & (-15.71, 2.11) \\
J_2 - J_4 & (-21.51, -3.69) \\
J_2 - J_5 & (-27.51, -9.69) \\
J_3 - J_4 & (-17.71, 3.11) \\
J_3 - J_5 & (-20.71, -2.89) \\
J_4 - J_5 & (-14.5, 2.5) \\
\end{array}
\]

If a confidence interval for the contrast \( J_1 + J_2 - (J_3 + J_4) \) is desired, \( K = 1 \), provided we are considering this contrast and nothing else. \( t_{\alpha/2} = 2.18 \) and

\[
\left( \frac{1}{r} \sum C_{h_i}^2 \right)^{1/2} = \left( \frac{1}{4.4} \right)^{1/2} = 1.
\]

Therefore, the critical constant is \((2.18)(3.66) = 7.98\) and the confidence interval is

\[
65.3 + 62.7 - 69.5 + 75.3 \pm 7.98 = (-24.78, 8.82).
\]

Notice that for pairwise differences in means, the Scheffé method gives the longest intervals and Tukey's method the shortest with the length of the Bonferroni intervals in between. Generally, Tukey's method will give the shortest confidence intervals for pairwise differences in means unless not all comparisons are of interest, in which case, the Bonferroni intervals may be shorter.

For the contrast \( J_1 + J_2 - (J_3 + J_4) \), the length of the Scheffé interval is shorter than the length of the Tukey interval. This is true in general. Note that for this single contrast, the Bonferroni method gives the shortest interval of the three. However, as the number of contrasts under consideration increases the critical
constant increases when you use the Bonferroni method. It is true that for a large number of contrasts, the Scheffe technique will give shorter intervals than will the Bonferroni method.

Section 5.2 Application of the Scheffe, Tukey, and Bonferroni techniques to a two-way classification design.

Next consider the following two-way design with replication: In the course of deciding on tactical uses for two rockets, an operation analyst requests the experiment reported in the following table. Three types of planes are used five times each with two types of rockets. The entries in the table represent coded evaluations of target destruction obtained in the test.

<table>
<thead>
<tr>
<th>Plane types</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>$\Sigma y_{ijk}$</th>
<th>$\bar{y}_{i..}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rocket types</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>64</td>
<td>4.27</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>74</td>
<td>4.93</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\Sigma y_{ijk}$ = 38 36 64

$\bar{y}_{..} = 4.6$

$\bar{y}_{ij} = 3.8 3.6 6.4$

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
## ANOVA

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among Rockets</td>
<td>1</td>
<td>3.33</td>
<td>33.3</td>
<td>1.12</td>
</tr>
<tr>
<td>Among Planes</td>
<td>2</td>
<td>48.8</td>
<td>24.4</td>
<td>8.22</td>
</tr>
<tr>
<td>Interaction</td>
<td>2</td>
<td>1.867</td>
<td>.933</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>24</td>
<td>71.2</td>
<td>2.967 = S²</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>125.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The appropriate model for this design is $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij} + e_{ijk}$, $i = 1, 2; j = 1, 2, 3; k = 1, 2, 3, 4, 5$, where $\Sigma \alpha_i = \Sigma \beta_j = \Sigma (\alpha \beta)_{ij} = (\alpha \beta)_{ij} = 0$. $S^2 = \frac{\sum (y_{ijk} - \bar{y}_{ij})^2}{24}$, $\hat{\alpha}_i = y_{1i.} - \bar{y}_{..}$, and $\hat{\beta}_j = y_{.j.} - \bar{y}_{..}$.

Since $1.12 < 4.26 = F_{1,24}^{.05}$, we would accept the hypothesis that there is no difference in the rocket types, but since $8.22 > 3.40 = F_{2,24}^{.05}$ we would reject the hypothesis that there is no difference among the planes and then consider differences in planes.

First consider a set of $95\%$ simultaneous confidence intervals for pairwise differences in plane types obtained using the Scheffé method. The probability statement applicable here is
\[ p \left( \sum_{j} c_{j} \hat{\beta}_{j} \in \sum_{i} c_{j} \hat{\beta}_{j} \pm \left( (c - 1) F_{c-1, rc(m-1)} \right)^{1/2} \cdot S \cdot \left( \sum_{i} \frac{c_{i}^{2}}{rm} \right)^{1/2} \right) = 1 - \alpha. \]

In this example \( r = 2, c = 3, \) and \( m = 5. \)
\( \hat{\beta}_{1} = 3.8 - 4.6 = -0.8, \hat{\beta}_{2} = 3.6 - 4.6 = -1.0, \hat{\beta}_{3} = 6.4 - 4.6 = 1.8 \)
\( s^{2} = 2.967; \) so \( s = 1.72. \) \( F_{2, 24}^{.05} = 3.40, 2F_{2, 24}^{.05} = 6.80, \) and
\( \sqrt{2F_{2, 24}^{.05}} = 2.61. \) For pairwise differences in means \( \left( \sum_{i} \frac{c_{i}^{2}}{rm} \right)^{1/2} = \sqrt{2} = .447. \) Therefore, the critical constant is
\( (1.72)(2.61)(.447) = 2.01. \) The simultaneous confidence intervals are
- plane differences \( \hat{\beta}_{j} - \hat{\beta}_{j}' \pm 2.01 \)
  - \( \beta_{1} - \beta_{2} \)
    \((-1.99, 2.21)\)
  - \( \beta_{1} - \beta_{3} \)
    \((-4.61, -0.59)^{*}\)
  - \( \beta_{2} - \beta_{3} \)
    \((-4.81, -0.79)^{*}\)

We conclude that plane III is more effective than either plane I or plane II, but that there is no significant difference in the effectiveness of planes I and II.

Next consider a set of 95 \( \% \) simultaneous confidence intervals for pairwise differences in column means (plane types) obtained using Tukey's method. The probability statement applicable here is
\[ p \left( \beta_{j} - \beta_{j}' \in \bar{y}_{j} - \bar{y}_{j}' \pm q_{c, rc(m-1)} \left( \frac{s}{\sqrt{rm}} \right) \cdot \left( \sum_{i} \frac{c_{i}^{2}}{rm} \right)^{1/2} \right) = 1 - \alpha. \]
where \( r = 2, c = 3, m = 5 \cdot q_{3,24}^{0.05} = 3.532, S = 1.72 \) and 
\[
\sqrt{rm} = \sqrt{10} = 3.16.
\]
The critical constant is \((3.532) \frac{1.72}{3.16} = 1.92\). The desired confidence intervals are

Plane differences \( \bar{y}.j - \bar{y}.j' \pm 1.92 \)

\[
\begin{align*}
\beta_1 - \beta_2 & \quad (-1.72, 2.12) \\
\beta_1 - \beta_3 & \quad (-4.52, -.68) \\
\beta_2 - \beta_3 & \quad (-4.72, -.88) 
\end{align*}
\]

Another alternative is simultaneous confidence intervals obtained from Bonferroni t-statistics. For \( K \) column contrasts the probability statement is

\[
p(\sum_{j=1}^{c} c_{hj} \beta_j \in \sum_{j=1}^{c} c_{hj} \bar{y}.j \pm t_{\alpha/2k} \cdot \frac{1}{\sqrt{rm}} \sum_{j=1}^{c} c_{hj}^2)^{1/2} \geq 1 - \alpha.
\]

For pairwise differences in planes, \( k = 3 \) and \((\sum_{j=1}^{c} c_{hj}) = \sqrt{2} = .447, r = 2, m = 5 \cdot t_{24}^{\alpha/2k} = 2.58, \) if \( k = 3 \). \( S = 1.72 \). Therefore, the critical constant is \((2.58)(1.72) (.447) = 1.984\) and the simultaneous confidence intervals are

Plane differences \( \bar{y}.j - \bar{y}.j' \pm 1.984 \)

\[
\begin{align*}
\beta_1 - \beta_2 & \quad (-1.784, 2.184) \\
\beta_1 - \beta_3 & \quad (-4.584, -.616) \\
\beta_2 - \beta_3 & \quad (-4.784, -.816) 
\end{align*}
\]

Observe that here, as in the first example, Tukey's
method gives us the shortest confidence intervals; whereas the Scheffé method gives the longest intervals for these pairwise mean differences. As before, the lengths of the Bonferroni intervals fall in between.

Section 5.3 Application of the Scheffé, Tukey, and Bonferroni techniques to one-way classification design.

Finally, consider the following one-way design: In a preliminary evaluation of three tranquilizing drugs, time limitation and the possibility of residual effects decreed that each subject receive only one drug. 18 psychiatric patients with similar diagnoses were rated with respect to anxiety on a seven point scale. Six were randomly assigned to each of the 3 drugs, and after several days each patient was blindly rated on the same scale. These changes in anxiety ratings were observed.

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug</td>
<td>2</td>
<td>21</td>
<td>10.5</td>
<td>12.1</td>
</tr>
<tr>
<td>Within Drugs</td>
<td>15</td>
<td>13</td>
<td>0.867=S²</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>34</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\frac{\sum_{j} y_{ij}}{j} = 9 \\
\frac{\bar{y}_j}{j} = 3.5 \\
\frac{\bar{y}_i}{i} = 1.5 \\
\frac{\bar{y}}{1} = 1.0
\]
The appropriate model is $y_{ij} = \mu_i + e_{ij}$, $i = 1, 2, 3$, $j = 1, 2, \ldots, 6$. $\hat{\mu}_i = \bar{y}_i$, $i = 1, 2, 3$ and $S^2 = \frac{\sum_{ij} (y_{ij} - \bar{y}_i)^2}{15}$.

Since $12.1 > 3.68 = F_{12,15}^{0.05}$, we would reject the hypothesis that all three drugs are equal in effect and to look for differences in drugs.

If a set of 95 o/o simultaneous confidence intervals for contrasts is desired using the Scheffè method, the appropriate probability statement is

$$p\left\{\sum_{i=1}^{r} c_i \mu_i = \sum_{i=1}^{r} c_i \bar{y}_i \pm \sqrt{(r-1)F_{r-1,N-r}^{0.05}} \cdot \left(\sum_{i=1}^{r} \frac{c_i^2}{n_i}\right)^{1/2}\right\} = 1 - \alpha.$$  

for all $c \in \mathbb{R}^r$.

In this problem, $r = 3$, $N - r = 15$, and $n_i = 6$, $i = 1, 2, 3$.

$\sqrt{2F_{2,15}^{0.05}} = \sqrt{2(3.68)} = 2.71$. $S^2 = .867$ and $5 = .931$. For pairwise differences in means $\left(\sum_{i=1}^{r} \frac{c_i^2}{n_i}\right)^{1/2} = (\frac{1}{6} + \frac{1}{6})^{1/2} = .574$. Therefore, the critical constant is $(2.71)(.931)$ ($(.574) = 1.45$, and the simultaneous confidence intervals are

Mean differences $\bar{y}_i - \bar{y}_1$, $\pm 1.45$

$\mu_1 - \mu_2$ $\quad (.55, 3.45)^*$

$\mu_1 - \mu_3$ $\quad (1.05, 3.95)^*$

$\mu_2 - \mu_3$ $\quad (-.95, 1.95)$

We can conclude that drug A is significantly better than either drug B or C, but that there is no significant difference between drugs B and C.
To obtain 95% simultaneous confidence intervals for pairwise differences in drug means using Tukey's method, we use the following probability statement:

\[
p\left( \sum c_i \mu_i \in \sum c_i \bar{y}_i \pm q_{\alpha, r, r(n-1)} \frac{S}{\sqrt{n}} \sum c_i \right) \text{ for all } c \in \mathbb{Z} = 1 - \alpha.
\]

Here, \( r = 3, n = 6, q_{3, 15}^{.05} = 3.674, S = .931 \) and \( \sqrt{6} = 2.45 \). The critical constant then is \( (3.674) \frac{.931}{2.45} = 1.396 \) and the simultaneous confidence intervals are

Mean differences \( \bar{y}_i - \bar{y}_j \pm 1.396 \)

\( \mu_1 - \mu_2 \) \( (.604, 3.396)* \)

\( \mu_1 - \mu_3 \) \( (1.104, 3.896)* \)

\( \mu_2 - \mu_3 \) \( (-.996, 1.896) \)

To obtain 95% simultaneous confidence intervals for pairwise differences in drug means using Bonferroni t-statistics we use the following statement:

\[
p\left( \sum c_i \mu_i \in \sum c_i \bar{y}_i \pm t_{\alpha/2k, N-r} \frac{1}{\sqrt{\sum \frac{c_i^2}{n_i}}} \right), h = 1, 2, \ldots, k \geq 1 - \alpha.
\]

In this case \( r = 3, k = 3, N - r = 15 \cdot t_{.05/2(6)} = 2.69, \)

\( S = .931 \) and \( \left( \sum \frac{c_i^2}{n_i} \right)^{1/2} = \sqrt{\frac{1}{3}} = .574 \). Therefore, the critical constant is 1.44 and the desired confidence intervals are

Mean differences \( \bar{y}_i - \bar{y}_j \pm 1.44 \)

\( \mu_1 - \mu_2 \) \( (.56, 3.44)* \)

\( \mu_1 - \mu_3 \) \( (1.06, 3.94)* \)

\( \mu_2 - \mu_3 \) \( (-.94, 1.94) \)
BIBLIOGRAPHY
BIBLIOGRAPHY

