CONFIDENCE INTERVALS FOR THE DIFFERENCES BETWEEN TREATMENT MEANS IN AN ANALYSIS OF VARIANCE

by

Robert G. Bonner

Thesis submitted to the Graduate Faculty of the Virginia Polytechnic Institute in candidacy for the degree of

MASTER OF SCIENCE in

STATISTICS

APPROVED:

Director of Graduate Studies

Dean of Applied Science and Business Administration

APPROVED:

Head of Department

Supervisor

May 1954

Blacksburg, Virginia
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>INTRODUCTION</td>
<td>3</td>
</tr>
<tr>
<td>II</td>
<td>REVIEW OF LITERATURE AND DISCUSSION</td>
<td>6</td>
</tr>
<tr>
<td>III</td>
<td>PROPOSED METHOD</td>
<td>13</td>
</tr>
<tr>
<td>IV</td>
<td>ILLUSTRATIVE EXAMPLE</td>
<td>18</td>
</tr>
<tr>
<td>V</td>
<td>PROPERTIES OF THE PROPOSED INTERVALS</td>
<td>23</td>
</tr>
<tr>
<td>VI</td>
<td>ALTERNATIVES FOR HANDLING COMPLEXITY OF NUISANCE PARAMETERS</td>
<td>26</td>
</tr>
<tr>
<td>VII</td>
<td>SUMMARY AND CONCLUSION</td>
<td>32</td>
</tr>
<tr>
<td>VIII</td>
<td>ACKNOWLEDGMENTS</td>
<td>33</td>
</tr>
<tr>
<td>IX</td>
<td>BIBLIOGRAPHY</td>
<td>34</td>
</tr>
<tr>
<td>X</td>
<td>VITA</td>
<td>36</td>
</tr>
</tbody>
</table>

## CHARTS

- **FIGURE 2.1**: COMPARISON OF SIGNIFICANCE TEST PROPOSED BY NEWMAN AND TEST SUGGESTED BY TUKEY'S CONFIDENCE INTERVALS
- **FIGURE 2.2**: COMPARISON OF REGIONS OBTAINED USING A 95 AND A 90.25 PER CENT LEVEL F TEST
- **FIGURE 3.1**: CONFIDENCE INTERVALS OBTAINED BY AN "INVERSION" OF THE MULTIPLE COMPARISONS TEST
- **FIGURE 6.1**: CONFIDENCE INTERVAL FOR $\delta_{1.2}$ WHICH IS AN AVERAGE WITH RESPECT TO A FIDUCIAL DISTRIBUTION OF $\delta_{12.3}$
I. INTRODUCTION

Several methods have been proposed for obtaining confidence intervals for the differences between treatment means in an analysis of variance. The purpose of this thesis is to present a method based upon the Multiple Comparisons Test (D. B. Duncan, 1951, 1952) which gives uniformly shorter intervals than previous procedures by using systems of joint confidence coefficients based on degrees of freedom.

The general problem is that of making inferences about differences among a set of treatment means like those illustrated in the following example taken from C. L. Davies (1959).

Table 1.1

<table>
<thead>
<tr>
<th>Sample Number of H-Acid</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Yields in grams of Naphthalene Black 12 B</td>
<td>1.40</td>
<td>1.90</td>
<td>1.50</td>
<td>1.40</td>
<td>1.55</td>
<td>1.45</td>
</tr>
<tr>
<td></td>
<td>1.50</td>
<td>1.95</td>
<td>1.55</td>
<td>1.45</td>
<td>1.55</td>
<td>1.45</td>
</tr>
<tr>
<td></td>
<td>1.55</td>
<td>1.58</td>
<td>1.56</td>
<td>1.55</td>
<td>1.60</td>
<td>1.56</td>
</tr>
<tr>
<td>Means</td>
<td>1.505</td>
<td>1.528</td>
<td>1.564</td>
<td>1.498</td>
<td>1.600</td>
<td>1.570</td>
</tr>
</tbody>
</table>

These data were obtained in an investigation of the differences in yield of Naphthalene Black 12 B arising from the use of six different H-acids. The aim of this thesis is to obtain confidence intervals for the differences such as those between the six mean yields in Table 1.1.
The analysis of variance obtained by the statistical evaluation of the results is presented in Table 1.2.

Table 1.2
Analysis of Variance

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Samples</td>
<td>5</td>
<td>56,357.5</td>
<td>11,271.5</td>
<td>4.596</td>
</tr>
<tr>
<td>Error</td>
<td>24</td>
<td>58,630.0</td>
<td>2,451.25</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>115,187.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The estimated standard error for a treatment mean is $\sqrt{24 \times 121.875} = 22.136$ and is based on 24 degrees of freedom.

The first impulse in answering this problem is to use Student's t-distribution for constructing confidence intervals for the differences among the observed means. That is, to use limits of the familiar form

\[(1.1) \quad d \pm \sqrt{t_{n_2}^2 s^2 /
2
}\]

where \(d = n_2 - n_3\) and \(n_1\) and \(n_2\) are any two means from the complete set, and \(t_{n_2}^0\) is the appropriate value from the \(t_{n_2}\) distribution.

Thus if 95 per cent intervals are required in the above example, \(t_{n_2}^0\) would be the five per cent significant value obtained from the \(t\)-distribution with \(n_2 = 24\) degrees of freedom. Individually, these intervals are all right in the sense that the probability that each
interval contains its corresponding parameter, the true difference, is 0.95. However, the probability that r such intervals will simultaneously contain the corresponding r true differences proves to be unduly low for r > 2. For example, if 95 per cent confidence intervals are determined for all differences among three means and among four means (in the case n_2 = ∞) the probabilities that all intervals will contain their true differences can readily be shown to be 0.8777 and 0.7966, respectively. These would generally be considered too low to be consistent with the requirement that the confidence coefficient for each interval be 0.95.

It will be useful in further discussion to introduce the term "joint confidence coefficient for p means" to denote the probability that the intervals determined for all differences among p means correctly contain their corresponding difference parameters. In these terms we may then say that the joint confidence coefficient for any subset of two means in the above procedure is 0.95 which is satisfactory. However, the joint confidence coefficients for p = 3 and p = 4 means are 0.8777 and 0.7966, respectively. In a case involving more means these coefficients would get rapidly lower as p increases. All coefficients except for p = 2 would be too low to be acceptable.
II. REVIEW OF LITERATURE AND DISCUSSION

We are given \( n \) treatment means \( m_1, m_2, \ldots, m_n \), which are normally distributed about \( n \) respective population means \( \mu_1, \mu_2, \ldots, \mu_n \) with common variance \( \sigma^2 \). We are also given an estimate \( s_m^2 \) of \( \sigma^2 \) which is obtained from an analysis of variance and has the usual distribution, namely \( n_2 s_m^2 / \sigma^2 \sim \chi^2_{n_2} \). The variance estimate \( s_m^2 \) is independent of \( m_1, m_2, \ldots, m_n \), and \( n_2 \) represents the error degrees of freedom in the analysis of variance.

For brevity and simplicity in most of our discussion we shall restrict ourselves to the special case when \( n = 3 \), \( n_2 = \infty \), and \( s_m^2 = \sigma^2 = 1 \).

The first procedure proposed for handling the stated problem was presented by J. W. Tukey (1951). Although several people had developed significance tests for the differences between treatment means, this was the first method explicitly proposed for obtaining confidence intervals of the type under consideration in this thesis.

Tukey proposed the 95 per cent interval, or 95 per cent allowance as he termed it,

\[
(2.1) \quad d \pm q_{p, n_2, 0.05} s_m,
\]

where \( d \) is the observed difference between two means as before, and \( q_{p, n_2, 0.05} \) is the five per cent significant value of the studentized range for \( p \) treatment means. By using the five per cent significant value, this method increases the joint confidence coefficient for the \( p \) means involved to 0.95.
Now we note that if Tukey's method were used for a
significance test it would give the rule:

\[ d \text{ is significant if } |d| \text{ exceeds } q_{p, n, 0.05} s_m \]

This test may be represented graphically for the case \( n = 3, \ n = \infty \)
\( s_m^2 = o_m^2 = 1 \), as in Figure 2.1. In this case \( q_{3, \infty, 0.05} = 3.32 \).

Sample and parameter spaces which are convenient for
illustrating the discussion have axes defined as follows:

\[
\begin{align*}
(2.2) \quad d_{1.2} &= \mu_1 - \mu_2, \\
(2.3) \quad d_{12.3} &= (\mu_1 + \mu_2 - 2\mu_3)/\sqrt{3}, \\
\delta_{1.2} &= \mu_1 - \mu_2, \\
\delta_{12.3} &= (\mu_1 + \mu_2 - 2\mu_3)/\sqrt{3}.
\end{align*}
\]

These are the axes used in Figure 2.1 and they will be used
throughout the thesis.

All points lying to the right of \( d = 3.32 \) represent
observed differences which lead to the decision \( \mu_1 > \mu_2 \), all
points lying to the left of \( d_{1.2} = -3.32 \) represent observed differences
which lead to the decision \( \mu_1 < \mu_2 \), and all points between these
lines represent non-significant differences.

The significance test suggested by Tukey's confidence
intervals may be compared in the same special case with a test
proposed by Newman (1939) and later, in more detail, by Keuls (1952).
This test called the Multiple Range Test states that any difference
\( d \) is significant if \( |d| > q_{2, \infty, 0.05} = 2.77 \), provided also that the
range exceeds \( q_{3, \infty, 0.05} = 3.32 \). The regions of this test are
represented by the dotted lines in Figure 2.1 and it can be seen to
be more powerful than Tukey's test.
NOTE:
HEXAGON IN BACKGROUND IS REGULAR AND MAY BE USED TO IDENTIFY THE BOUNDARIES OTHERWISE UNSPECIFIED.

FIGURE 2.1 COMPARISON OF SIGNIFICANCE TEST PROPOSED BY NEWMAN AND TEST SUGGESTED BY TUKEY'S CONFIDENCE INTERVALS
Newman's procedure will find a significant difference whenever Tukey's will and on other occasions also. This suggests that confidence intervals obtained as a corollary of Newman's test would be shorter on the average than those employed by Tukey. It is noteworthy that Tukey (1953) has moved towards Newman's approach as far as significance tests are concerned, but has not changed his method for obtaining confidence intervals. Confidence intervals obtained from the Newman - Keuls test would have the same joint confidence coefficient for the three means as Tukey's method, that is, 0.95 in the given example. The difference in the two methods lies in the confidence coefficients which each one gives for the means considered in pairs. In Tukey's method this coefficient is

\[ P\left\{ t < q_2/\sqrt{2} \right\} = P\left\{ t < 2.35 \right\} = 0.9812 \]

where \( t \) has the \( N(0, 1) \) distribution. In the Newman test this value would lie between 0.9612 and 0.95, the actual value depending on a nuisance parameter which will be discussed in more detail.

An \( F \) test may be used instead of the preliminary range test in Newman's procedure. The regions for a procedure obtained in this way would be as shown in Figure 2.2 for the same special case as before. In words the test would be as follows: If \( |d_{1.2}| > \sqrt{2} t_{n_2} s_m, \) \( d_{1.2} \) is significant, provided also that an \( F \) test for the variance between \( m_1, m_2, m_3 \) is significant. When in place of a preliminary range test a preliminary \( F \) test is used, it is legitimate, if desired, to apply the subsequent t-tests not only to differences between single means but to all comparisons of the form.
FIGURE 2.2 COMPARISON OF REGIONS OBTAINED USING A 95 AND A 90.25 PERCENT LEVEL F TEST

NOTE:
CIRCLES IN BACKGROUND MAY BE USED TO IDENTIFY BOUNDARIES OTHERWISE UNSPECIFIED
(2.4) \[ C = \sum_{i=1}^{p} k_i m_i \quad \text{where} \quad \sum_{i=1}^{p} k_i = 0. \]

Now, in a five per cent level test, Newman used a five per cent level significant value for \( q_3 \) and for \( q_2 \). It can be argued, however, that if a five per cent level value is appropriate for \( q_2 \), a 9.75 per cent value would be appropriate for \( q_3 \). For, in any combination of three treatments, we could have chosen two independent comparisons \text{a priori}. Had this been done it would have been legitimate to have applied five per cent level independent tests to each comparison. Under these circumstances, the probability of rejecting the hypothesis that all three treatment means are equal would have been 9.75 per cent. A discussion of this point is given by Duncan (1951, 1952, 1953).

This principle is also relevant to the multiple F test procedure. Thus we would use a 9.75 level F test getting the changes shown by the dotted lines in Figure 2.2. The use of five per cent level t-tests preceded by a 9.75 per cent level F test in this way gives the Multiple Comparisons Test for the case \( n = 3, n_2 = \infty \), and \( \sigma_m^2 = \sigma_m^2 = 1. \)

This is the test we will "invent", so to speak, for getting the confidence intervals to be proposed in this thesis.

In the same way as the multiple F test allows one to test general comparisons of the form (2.4), so will the proposed confidence interval procedure give simultaneous confidence intervals for all comparisons of the form

(2.5) \[ \theta = \sum_{i=1}^{p} k_i m_i \quad \text{where} \quad E(\theta) = \sum_{i=1}^{p} k_i m_i \quad \text{and} \quad \sum_{i=1}^{p} k_i = 0. \]
As a rule differences between two means will be of most interest. Because this is true and also for simplicity most of the discussion will be limited to the estimation of intervals of this kind. An important feature of these confidence intervals is derived from the Multiple Comparisons Test. The joint confidence coefficient for any p means is at least \((1 - \alpha)^p - 1 = \gamma^p - 1\) where \(\gamma\) is the confidence coefficient desired for each difference considered individually.

Recently, Henry Scheffé (1953) has published a confidence interval procedure which is like Tukey's in that the joint confidence coefficient for p means is kept as high as \(\gamma\). Like the proposed procedure it has much in common with the Multiple Comparisons Test in that it is based on F tests.
III. PROPOSED METHOD

As we have said, the aim of this thesis is to obtain a procedure for placing confidence intervals on the differences between treatment means by an "inversion", so to speak, of the Multiple Comparisons Test.

We shall first consider the confidence intervals obtained in this way for the case of three treatments, that is, for the case \( n = 3 \). Figure 3.1 shows the confidence interval for the difference \( \delta_{1.2} \) in the previously defined parameter space with axes \( \delta_{1.2} \) and \( \delta_{12.3} \), for the special case \( n = 3 \), \( n_2 = \infty \), \( s_m^2 = \sigma_m^2 = 1 \), and \( \alpha = .05 \).

Given a sample point \( (d_{1.2}, d_{12.3}) \), the confidence interval for \( \delta_{1.2} \) is the larger of the intervals given by the following limits:

\[
\begin{align*}
(3.1) & \quad d_{1.2} \pm s_m^2 \sqrt{2} t_{n_2}^\alpha, \\
(3.2) & \quad d_{1.2} \pm \sqrt{2(2s_m^2)} F_{2,n_2}^\alpha - (d_{12.3} - \delta_{12.3})^2, \\
\end{align*}
\]

where \( t_{n_2}^\alpha \) is the 5 per cent level value of \( t_{n_2} \) as before and \( F_{2,n_2}^\alpha \) is the (0.95) 100 percentile of the \( F \) distribution with 2 and \( n_2 \) degrees of freedom. (In general, if \( \alpha \) is the level of \( t_{n_2} \) value in (3.1), the \( F \) value in (3.2) is \( (1 - \alpha)^2 \) 100 percentile.)

The interval (3.1) is readily recognized as the usual form of the confidence interval for a difference between two means obtained from Student's \( t \)-distribution. This interval is represented by the parallel lines in Figure 3.1.

The interval (3.2) is obtained from the \( F \) distribution with 2 and \( n_2 \) degrees of freedom. Starting from the evident relationship
NOTES:

SPECIAL CASE

\[ n = 3 \]
\[ n_2 = \infty \]
\[ \sigma^2 = 1 \]
\[ m = 1 \]
\[ \alpha = 95 \text{ PER CENT} \]

NOTE:

CIRCLE IN BACKGROUND MAY BE USED TO IDENTIFY BOUNDARIES OTHERWISE UNSPECIFIED

FIGURE 3.1 CONFIDENCE INTERVALS OBTAINED BY AN "INVERSION" OF THE MULTIPLE COMPARISONS TEST
(3.3) \[ \frac{(d_{1.2} - \delta_{1.2})^2 + (d_{12.3} - \delta_{12.3})^2}{2 s_m^2 F_{2, n_2}} \]

cross-multiplying, rearranging terms, and substituting \( F_{2, n_2}^0 \) for \( F_{2, n_2} \) we get

(3.4) \[ (d_{1.2} - \delta_{1.2})^2 < 2(2s_m^2) F_{2, n_2}^0 - (d_{12.3} - \delta_{12.3})^2, \]

where \( F_{2, n_2}^0 \) as before is the \((1 - .05)\) 100 percentile of the F distribution with 2 and \( n_2 \) degrees of freedom. Taking the square root of both sides of (3.4) we find the confidence interval (3.2) for \( \delta_{1.2} \).

If the interval (3.1) were used entirely, the joint confidence coefficient for the three means, that is the probability that the intervals will correctly cover the three differences \( \delta_{1.2} \), \( \delta_{1.3} \) and \( \delta_{2.3} \) simultaneously, would be too low. The purpose of extending the central part of the interval to the limits (3.2) is to increase this coefficient to an acceptable value. This will be discussed after considering the general case.

For brevity and simplicity in presentation, it is advisable to introduce two symbols. The first, \( D_s \), shall denote deviations such as those in (3.4); that is \( D_s = d_s - \delta_s \) where \( s \) is the appropriate subscript. The second, \( R_{p, n_2}^0 \), shall denote the value \((p - 1) 2s_m^2 F_{p - 1, n_2}^0 \) such as that in (3.2). This is considered advisable because the value of \((p - 1) F_{p - 1, n_2}^0 \) is most easily obtained from Table I in Duncan (1951) Significant Ranges \( R_{p, n_2}^0 \) for a 5 per cent Level Multiple Comparisons Test. That is,
\[(3.5) \quad \begin{align*}
R^2_{\theta \phi} &= R^2_{\phi \theta} \quad s^2 = (p-1) \frac{2s^2_m P^0}{p - 1, n_2} \quad \text{and} \\
R^2_{\phi \theta} &= R^2_{\theta \phi} \quad s^2 = \sqrt{\frac{(p-1) 2s^2_m P^0}{p - 1, n_2}}
\end{align*}\]

In the general case, the proposed confidence interval for \( \delta_{1,2} \) is the largest interval given by the following limits:

\[(3.6) \quad d_{1,2} = s_m \sqrt{\frac{2}{n_2}} + t^0_{n_2}\]

\[(3.7) \quad d_{1,2} = \sqrt{\frac{R^2_{\phi \theta}}{n_2}} - D^2_{1,2, i}; \quad i = 3, h, \ldots n.
\]

(There are \( n-2G1 \) limits of this form.)

\[(3.8) \quad d_{1,2} = \sqrt{\frac{R^2_{\phi \theta}}{n_2}} - D^2_{1,2, i}; \quad i, j = 3, h, 5, \ldots n; \quad i \neq j.
\]

(There are \( n-2G2 \) limits of this form.)

\[(3.9) \quad d_{1,2} = \sqrt{\frac{R^2_{\phi \theta}}{n_2}} - D^2_{1,2, i j k}; \quad i, j, k = 3, h,
\]

(There are \( n-2G3 \) limits of this form.)

\[(3.10) \quad d_{1,2} = \sqrt{\frac{R^2_{\phi \theta}}{n_2}} - D^2_{1,2, i j k \ldots n}; \quad i, j, k = 3, h,
\]

(There is only \( n-2Gn-2 = 1 \) limit of this form.)

The purpose of these intervals is similar to that of the interval \( (3.2) \) in the case \( n = 3 \). Their inclusion increases each joint confidence coefficient for \( p \) means to \((.95)^P - 1\) for all values of \( p \), \( p = 2, 3, \ldots, n \).

These intervals are complex in that the interval for any difference, say \( \delta_{1,2} \) is dependent on the values of nuisance parameters.
like $\delta_{12.ij}$, $\delta_{1.j}$, and others in the limits given above.

Three alternatives are being considered for dealing with this complexity.

The first alternative is to accept the most conservative interval, that is, to accept the largest interval that can be obtained for all possible variations of the nuisance parameters. This would be given by replacing any nuisance parameter $\delta$ by the corresponding observed difference $d$, that is by putting $D = d - \delta = 0$. The longest interval for $\delta_{1.2}$ would then be the limit given by (3.10) which equals

\[
(3.11) \quad d_{1.2} \pm \sqrt{\frac{R^0}{n_1 n_2}} = d_{1.2} \pm R^0_{n_1 n_2}.
\]

It should be noted that this method of dealing with the nuisance parameters is an implicit feature in the allowances suggested by Tukey (1951).

Other approaches for dealing with the complexity will be discussed in a later section.
IV. ILLUSTRATIVE EXAMPLE

To illustrate the previous discussion, we shall place confidence intervals on the differences between the treatment means presented in Table 1.1.

The means for each of the six treatments are

<table>
<thead>
<tr>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1505</td>
<td>1528</td>
<td>1564</td>
<td>1493</td>
<td>1600</td>
<td>1470</td>
</tr>
</tbody>
</table>

The standard error for a treatment mean is 22.136 and is based on 24 degrees of freedom.

The proposed confidence interval for the difference \( \delta_{1.2} \) between the two means 1 and 2 is the largest interval given by the following limits:

\[
(l.1)
\delta_{1.2} \pm \frac{s}{m} \sqrt{\frac{2}{n_1}}
\]

which reduce to

\[
23 \pm 6.6, \quad 23 \pm 6.7
\]

and the following reduced in the same way:

\[
23 \pm \sqrt{5036.74 - (54.85 - \delta_{12.3})^2}
\]

\[
23 \pm \sqrt{5036.74 - (21.36 - \delta_{12.4})^2}
\]

\[
\ldots \ldots \ldots \ldots \ldots \ldots
\]

\[
23 \pm \sqrt{5036.74 - (53.70 - \delta_{12.6})^2}
\]
\[
23 \pm \sqrt{5816.13 - (-20.51 - \delta_{12.3h}^2) - (66 - \delta_{3.6h}^2)}
\]
\[
23 \pm \sqrt{5816.13 - (-92.64 - \delta_{12.35}^2) - (-36 - \delta_{3.5}^2)}
\]
\[
23 \pm \sqrt{5816.13 - (-26.17 - \delta_{12.56}^2) - (130 - \delta_{5.6}^2)}
\]
\[
23 \pm \sqrt{6617.82 - (-58.09 - \delta_{12.3h5}^2) - (17.32 - \delta_{3.45}^2) - (-102 - \delta_{4.5}^2)}
\]
\[
23 \pm \sqrt{6617.82 - (9.04 - \delta_{12.3h6}^2) - (92.38 - \delta_{3.4h6}^2) - (28 - \delta_{4.6}^2)}
\]
\[
23 \pm \sqrt{6617.82 - (-9.55 - \delta_{12.4h56}^2) - (42.73 - \delta_{4.56}^2) - (130 - \delta_{5.6}^2)}
\]

\((4.2)\)
\[
23 \pm \sqrt{7361.84 - \Sigma D^2} \quad \text{where} \quad \Sigma D^2 = \left(-26.14 - \delta_{12.3h56}^2 \right)^2 + \left(-5.66 - \delta_{3.4h56}^2 \right)^2 + (66 - \delta_{3.4}^2)^2 + (130 - \delta_{5.6}^2)^2.
\]

The first term in each square root, \(R^2_{p,n} \), is most easily evaluated from the relation

\[(4.3)\]
\[
R^0_{p,n} = R^0_{p,n} s_m,
\]

where the quantities \(R^0_{p,n} = \sqrt{(p - 1) \cdot 2 F(\text{p - 1),n}_2}
\]

have been tabulated by Duncan (1951, Table I). For example, under the first square root

\[
R^0_{p,n} = \sqrt{(R^0_{3,2h4} \times s_m)^2} = \sqrt{(3.306 \times 22.136)^2} = \sqrt{5036.74} = 70.970.
\]
The numerical value in each squared deviation is the estimate of the parameter subtracted from it. In calculating the estimate of the parameter, the means before the point have positive coefficients while those after the point have negative coefficients. The sum of the coefficients must equal zero. The above term is divided by the square root of the sum of squares of the coefficients divided by two. For example

\[
(d_{12.345}) = \frac{3(1505) + 3(1528) - 2(1564) - 2(1498) - 2(1600)}{\sqrt{30/2}}
\]

\[
d_{12.345} = \frac{-225}{\sqrt{15}} = -58.09
\]

If the device of replacing the nuisance parameters \( \delta \) by the observed difference \( d \) is adopted, the longest interval for \( \delta_{1.2} \) is \((k.2)\) and all that is necessary is to calculate:

\[
(k.5) \quad \frac{23 \pm \sqrt{\frac{R^2}{n-1,n_2}}}{n-1,n_2}, \quad \text{or} \quad 23 \pm \frac{R^1}{n-1,n_2} \quad \text{is} = 23 \pm 85.80.
\]

Similarly, the interval for any difference \( \delta_{i,j} \) is given by the limit

\[
(k.6) \quad d_{i,j} \pm 85.80
\]

The confidence intervals for all 602 differences among the treatment means in this illustration are presented in Table h.1.
Table 4.1
Confidence Intervals for the Differences Among the Six Treatment Means

<table>
<thead>
<tr>
<th>Difference</th>
<th>Confidence Interval</th>
<th>Difference</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower Limit</td>
<td>Upper Limit</td>
<td></td>
</tr>
<tr>
<td>(d_{1.2})</td>
<td>-108.80</td>
<td>62.80</td>
<td>(d_{2.6})</td>
</tr>
<tr>
<td>(d_{1.3})</td>
<td>-26.80</td>
<td>144.80</td>
<td>(d_{3.4})</td>
</tr>
<tr>
<td>(d_{1.4})</td>
<td>-92.80</td>
<td>78.80</td>
<td>(d_{3.5})</td>
</tr>
<tr>
<td>(d_{1.5})</td>
<td>-180.80</td>
<td>-9.20</td>
<td>(d_{3.6})</td>
</tr>
<tr>
<td>(d_{1.6})</td>
<td>-50.80</td>
<td>120.80</td>
<td>(d_{4.5})</td>
</tr>
<tr>
<td>(d_{2.3})</td>
<td>-121.80</td>
<td>49.80</td>
<td>(d_{4.6})</td>
</tr>
<tr>
<td>(d_{2.4})</td>
<td>-55.80</td>
<td>115.80</td>
<td>(d_{5.6})</td>
</tr>
</tbody>
</table>

By employing Tukey's procedure for obtaining confidence intervals which are of the form

\[(4.7) \quad d_{i,j} \pm q_{p,n_2} \frac{s_m}{\sqrt{m}}\]

where \(q_{p,n_2}\) in this illustration is the 5 per cent significant value of the studentized range for 6 means based on 21 degrees of freedom, we get

\[(4.8) \quad d_{i,j} \pm q_{p,n_2} (4.38) (22.136)\]

\[d_{i,j} \pm 96.96\]
By employing this procedure, the joint confidence coefficient for
the differences between any two means of a six mean set is increased
to 0.95, but the width of the interval obtained is 193.92 as compared
to 171.60 in the proposed method.

The intervals given by the proposed method are not only
shorter but since they are based on the F distribution they can be
applied to all contrasts of the form (2.5), as well as to differences
between single means. The basic reason for these improvements is that
in the proposed procedure the joint confidence coefficients for p means
have lower limits of \((.95)^p - 1\), whereas in the other procedure
(Tukey's allowances) the lower limits are (.95).
V. PROPERTIES OF THE PROPOSED INTERVALS

The proposed confidence intervals given above have the following properties:

(1) the probability that the interval will cover the parameter $\delta_{ij}$ is greater than or equal to 0.95, and

(2) if the same procedure is applied simultaneously to each of the $p\times 2$ differences among the $p$ means, the probability that all confidence intervals will cover the parameters correctly is $(0.95)^p - 1$.

In other words, the confidence coefficient for the difference between any two means is 0.95 and the joint confidence coefficient for all intervals for any subset of $p$ means is $(0.95)^p - 1$.

The proof of these properties in the case $n = 3$ is fairly evident from the discussion of that case in Section III. Here we shall deal with the case $n = 4$. The proof of the general case follows readily in the same way.

In this case, the confidence interval for $\delta_{1.2}$ is the largest interval given by the following limits:

\begin{align*}
(5.1) & \quad d_{1.2} \pm s_m \sqrt{2} t_{n_2}^0 \\
(5.2) & \quad d_{1.2} \pm \sqrt{ \frac{\hat{\tau}^0_{p-3}}{3_p n_2} - D_{12,3}^2 } \\
(5.3) & \quad d_{1.2} \pm \sqrt{ \frac{\hat{\tau}^0_{p-3}}{3_p n_2} - D_{12,3}^2 } \\
(5.4) & \quad d_{1.2} \pm \sqrt{ \frac{\hat{\tau}^0_{p-3}}{n_1 n_2} - D_{12,3}^2 - D_{3,4}^2 } .
\end{align*}
Similarly, the intervals for $\delta_{1.3}$, $\delta_{1.4}$, $\delta_{2.3}$, $\delta_{2.4}$, and $\delta_{3.4}$ are the largest intervals in each of five similar sets of four intervals each.

Now if our only objective were to provide a confidence interval for each difference $\delta_{i.j}$ with a confidence coefficient of 0.95, the interval we would get would be that of the limit

$$d_{i.j} \pm s_m \sqrt{\frac{2}{n_2}} t^o_{n_2}$$

Since this is a minimum interval for $\delta_{i.j}$ in the proposed method, the confidence coefficient for the difference between any two means is at least 0.95.

In the same way if our sole objective were to provide a method having a joint confidence coefficient of 0.9025 for any three-mean subset, we would proceed as follows. For any such subset $\mu_i$, $\mu_j$, $\mu_k$, we can make the probability statement

$$p \left\{ \frac{D_{i.j}^2 + D_{ij.k}^2}{2} \leq \frac{s_m^2}{m^2} \right\} = 0.9025 .$$

Now the inequality in this statement is always satisfied by the intervals given simultaneously for $\delta_{i.j}$, $\delta_{i.k}$, and $\delta_{j.k}$.

In the case of $\delta_{i.j}$ a minimum length for the interval is

$$d_{i.j} - \frac{\sqrt{R_{i,j}^2}}{\sqrt{n_2}} - D_{ij.k} < \delta_{i,j} < d_{i.j} + \frac{\sqrt{R_{i,j}^2}}{\sqrt{n_2}} + D_{ij.k} .$$

If $\delta_{i.j}$ be given any of the values in this interval it is clear that the inequality (5.6) is satisfied.

In the case of $\delta_{i.k}$ a minimum length for the interval is
\[(5.8) \quad d_{i,k} - \sqrt{\frac{\nu_2^2}{2\lambda_3n_2} - \mu_{k,j}^2} < \delta_{i,k} < d_{i,k} + \sqrt{\frac{\nu_2^2}{2\lambda_3n_2} - \mu_{k,j}^2},\]

which reduces to the inequality

\[(5.9) \quad \frac{d_{i,k}^2 + d_{k,j}^2}{2} / 2a_2^2 \leq \nu_2^2.\]

In the same manner it can be shown that \(\delta_{j,k}\) has a minimum interval which implies the satisfaction of (5.6). Thus the use of these intervals insures a joint confidence coefficient of at least 0.9025 for the means \(\mu_i, \mu_j, \mu_k\). In the proposed procedure the interval given by the limit (5.7) or its equivalent is the minimum interval for the difference between any two of the three means under consideration.

If, in the case of a four-mean subset, our sole aim were to provide a method having a joint confidence coefficient of 0.8574, we would proceed as follows. We can make the probability statement

\[(5.10) \quad P \left\{ \frac{D_{i,j}^2 + D_{i,j,k}^2 + D_{k,l}^2}{2} / 2a_2^2 \leq \nu_2^2 \right\} = 0.8574.\]

Thus the probability that the interval given by the limit

\[(5.11) \quad d_{i,j} + \sqrt{\frac{\nu_2^2}{2\lambda_3n_2} - D_{i,j,k}^2 - D_{k,l}^2}\]

will cover all parameters \(\delta_{i,j}\) is 0.8574. The use of the interval (5.11) insures a 0.8574 joint confidence coefficient for the four means \(\mu_i, \mu_j, \mu_k, \mu_l\).

In the proposed procedure the interval given by the limit (5.11) or its equivalent is the minimum interval for the difference between any two of the four means under consideration.
VI. ALTERNATIVES FOR HANDLING
COMPLEXITY OF NUISANCE PARAMETERS

As noted previously the proposed intervals for any
difference parameter $\delta_{i,j}$ is dependent on the values of nuisance
parameters like $\delta_{ij,k}$, $\delta_{ij,kl}$, $\delta_{ij,klm}$. One method for handling
nuisance parameters is to accept the most conservative interval,
that is, to accept the longest interval that can be obtained for all
possible variations of the nuisance parameters. This would be given
by replacing each nuisance parameter $\delta$ by the corresponding observed
difference d. This method was discussed and illustrated in Section
III. Two other alternatives for dealing with nuisance parameters can
be entertained.

One alternative is to express the intervals in terms of the
nuisance parameters and use subsequent or a priori knowledge to place
limits on the nuisance parameters. Then finally choose the longest
interval that can be obtained from variations of these parameters over
their restricted ranges.

Using this method for the case of three treatments, that is,
case n = 3, the confidence intervals for $\delta_{1,2}$ would be obtained as
follows. First, we recall that the required interval is the largest
given by the following limits:

\[ d_{1,2} \pm s_m \sqrt{\frac{2}{2n}} t_{n_2}^0, \]

\[ d_{1,2} \pm \sqrt{2(2s_m^2) F_{2,n_2}^0 - (d_{12,3} - \delta_{12,3})^2}, \]

(6.1)

(6.2)
where \( d_{1.2} \) and \( d_{12.3} \) are the observed differences, and \( \delta_{12.3} \) is the nuisance parameter to be removed. We note that the interval to be obtained will depend largely on the square of the deviation

\[
D_{12.3}^2 = d_{12.3}^2 - \delta_{12.3}^2.
\]

If, by a priori or subsequent knowledge, it seems reasonable to assume that the nuisance parameter \( \delta_{12.3} \) lies between \( \delta_{12.3} \) and \( \delta_{12.3}^0 \), choose \( \delta_{12.3}^0 \) between these limits so that the squared deviation \( D_{12.3}^2 \) will be minimized. Then choose the longer of the two intervals given by the following limits:

\[
d_{1.2} + s \sqrt{2} t^0_{n_2},
\]

\[
d_{1.2} + \sqrt{\frac{R^2}{3n_2} - D_{12.3}^2}.
\]

The third alternative would be to obtain an interval for the parameter \( \delta_{1.2} \) which is an average with respect to a fiducial distribution of the nuisance parameter. Treatment of the special case of three treatments, case \( n = 3, n_2 = \infty, \delta_m^2 = 1, \alpha = 5 \) per cent, will serve to illustrate the basic principles and calculations necessary.

From previous work the confidence interval for \( \delta_{1.2} \) is the larger interval given by the two limits:

\[
d_{1.2} + L_1 \text{ where } L_1 = s_m \sqrt{2} t^0_{n_2},
\]

\[
d_{1.2} + L_2 \text{ where } L_2 = \sqrt{\frac{R^2}{3n_2} - D_{12.3}^2}.
\]

In the special case we are considering
\[ L_1 = \sqrt{2} \ t_{\infty,0.05} = 1.881 \times 1.260 = 2.7718, \]
\[ L_2 = \sqrt{R_{3,\infty,0.05}^2 - D_{12.3}^2} = \sqrt{9.308 - D_{12.3}^2}, \]
where \( D_{12.3} = d_{12.3} - \delta_{12.3} \) as defined.

If we write \( d_{12.2} \pm L \) for the desired interval then
\[ L = L_1 \quad \text{when} \]
\[ \sqrt{R_{3,\infty,0.05}^2 - (d_{12.3} - \delta_{12.3})^2} < \sqrt{2} \ t_{\infty,0.05} \]
and \( L = L_2 \) when this inequality is reversed.

Our problem lies in the fact that \( \delta_{12.3} \) is unknown and \( L_2 \)
cannot be determined. The fiducial approach for getting around this
difficulty is equivalent to noting first that although \( \delta_{12.3} \) is
unknown, the distribution of \( u = d_{12.3} - \delta_{12.3} \) is known. In this
special case \( u \) is distributed normally with mean zero and unit variance.

Next we note that the inequality (6.7) is satisfied when
\[ \sqrt{9.308 - u^2} < \sqrt{2} \ t_{\infty,0.05} \]
that is, when \( |u| < u_o = 0.90150 \)

The required average value \( L_A \) of \( L \), averaged over the
fiducial distribution of \( \delta_{12.3} \) is seen to be given by an average of
\( L \) averaged over the distribution of \( u \), that is
\[ L_A = \int_{-\infty}^{\infty} L f(u) \, du = 2 \int_0^\infty L f(u) \, du \]
\[ = 2 I + 2 L_1 F[u > u_0], \]
where \( I = \int_{-\infty}^{u_0} L_2 f(u) \, du \).
Unfortunately, it is necessary to evaluate the integral I by a numerical method. Table 6.1 shows the work involved in using a six-panel Lagrange multiplier method for which the coefficients are given by Fisher and Yates in Table XXIV, Calculation of Integrals from Equally Spaced Ordinates.

Table 6.1
Computations Necessary for the Numerical Integration of Integral I

<table>
<thead>
<tr>
<th>u</th>
<th>f(u) *</th>
<th>L₂</th>
<th>g(u) = L₂f(u)</th>
<th>Lagrange Multipliers φ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.396,94,2,3</td>
<td>3.0509</td>
<td>1.217,13</td>
<td>21.1</td>
</tr>
<tr>
<td>.150,25</td>
<td>0.391,4,79,1</td>
<td>3.0472</td>
<td>1.202,06</td>
<td>216</td>
</tr>
<tr>
<td>.300,50</td>
<td>0.381,387,2</td>
<td>3.0361</td>
<td>1.157,93</td>
<td>27</td>
</tr>
<tr>
<td>.450,75</td>
<td>0.360,525,8</td>
<td>3.0174</td>
<td>1.087,85</td>
<td>272</td>
</tr>
<tr>
<td>.601,00</td>
<td>0.333,222,6</td>
<td>2.9911</td>
<td>0.996,70</td>
<td>27</td>
</tr>
<tr>
<td>.751,25</td>
<td>0.301,13,6</td>
<td>2.9570</td>
<td>0.890,66</td>
<td>216</td>
</tr>
<tr>
<td>.901,50</td>
<td>0.266,081,6</td>
<td>2.9117</td>
<td>0.775,55</td>
<td>21.1</td>
</tr>
</tbody>
</table>

* Ordinates of the normal distribution.

\[
\frac{\sum \phi g(u)}{\sum \phi} = \frac{687.754}{340} = 1.966,850
\]

I is now obtained by multiplying 1.966,850 by the length of the interval 0.90150. Therefore

\[
(6.10) \quad I = \int_{0}^{u} L_2 F(u) \, du = 0.90150 \times 1.966,850 = 0.952,75
\]
Now, $P(u > u_o)$ may be found in Table II, Pearson (1931) as

$$1 - 0.8160 = 0.184.$$ The length of $L_1$ is $\sqrt{2} u_{.05} = 1.4142 x 1.960 = 2.7718$ and the product $L_1 P(u > 0.9015) = (2.7718) (0.184) = 0.510.01$. Then the expected value $L_A$ of the length of the confidence interval is

$$(6.11) \quad L_A = 2I + 2L_1 P(u > u_o) = 2 \times 0.95275 + 2 \times 0.510.01$$

$$= 2.92552$$

when $n = 3$, $n_2 = \infty$, $\sigma_m = 1$, and $\alpha = 5$ per cent.

Figure 6.1 shows the limits $L_1 = 2.77$, $L_2 = 3.05$ when $D_{12,3} = 0$, and the average value $L_A = 2.93$ of $L$ averaged over the distribution of $u$. This may be compared with Figure 3.1 where the interval was given by the limit $d_{1,2} \pm 3.05$.

In more general cases, similar fiducial averages could be derived but would unfortunately involve more complex numerical integrations. The limits of the multivariate integrals required would not be simple to work with.

More thought needs to be given to the value of this approach before one would embark on making the tables required for its use in general practice.
NOTE:
CIRCLE IN BACKGROUND MAY BE USED TO IDENTIFY BOUNDARIES OTHERWISE UNSPECIFIED

FIGURE 6.1 CONFIDENCE INTERVAL FOR $\delta_{1.2}$ WHICH IS AN AVERAGE WITH RESPECT TO A FIDUCIAL DISTRIBUTION OF $\delta_{12.3}$

SPECIAL CASE

$n = 3$

$n_2 = \infty$

$m = 1$

$\alpha = 5$ PER CENT
VII. SUMMARY AND CONCLUSIONS

A method has been proposed for obtaining confidence intervals for the differences between treatment means in an analysis of variance. The intervals have the following properties:

(1) the probability that the interval will cover the parameter is greater than or equal to \((1 - \alpha)\), and

(2) if the same procedure is applied simultaneously to each of the \(p^2\) differences among \(p\) means, the probability that all confidence intervals will cover the parameters correctly is at least \((1 - \alpha)^p - 1\).

The same properties hold if the procedure is simultaneously applied to special linear comparisons among the means as well as to differences between single means. The intervals are complex in that the limits are dependent on the values of nuisance parameters. Three alternatives for handling these nuisance parameters are discussed, and one is preferred for use in practice.
VIII. ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to Professor David B. Duncan for his guidance during the course of this work. The results owe much to his many ideas and suggestions.

The author would also like to acknowledge with deep thanks the willing assistance of

in preparing the final copies for presentation.
IX. BIBLIOGRAPHY


The vita has been removed from the scanned document