ANALYSIS AND COMPUTER PROGRAMMING OF
DUNCAN'S NEW MULTIPLE RANGE TEST

APPROVED:

Earl W. Hooker
Major Professor

Carl T. York
Minor Professor

Darin Blank
Dean of the School of Education

Robert B. Trouche
Dean of the Graduate School
ANALYSIS AND COMPUTER PROGRAMMING OF DUNCAN'S NEW MULTIPLE RANGE TEST

THESIS

Presented to the Graduate Council of the North Texas State University in Partial Fulfillment of the Requirements For the Degree of

MASTER OF SCIENCE

By

Walter Wheeler, B.S.
Denton, Texas
June, 1965
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF ILLUSTRATIONS</td>
<td>v</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Purpose</td>
<td></td>
</tr>
<tr>
<td>Historical Background</td>
<td></td>
</tr>
<tr>
<td>II. NEW MULTIPLE RANGE TEST</td>
<td>13</td>
</tr>
<tr>
<td>Assumptions</td>
<td></td>
</tr>
<tr>
<td>Computer Program</td>
<td></td>
</tr>
<tr>
<td>Program Explanation</td>
<td></td>
</tr>
<tr>
<td>Data Preparation</td>
<td></td>
</tr>
<tr>
<td>Alternate Program</td>
<td></td>
</tr>
<tr>
<td>Analysis of Data</td>
<td></td>
</tr>
<tr>
<td>Computation Example</td>
<td></td>
</tr>
<tr>
<td>III. SUMMARY</td>
<td>44</td>
</tr>
<tr>
<td>APPENDIX</td>
<td>49</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>56</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table                                         Page

I. Data Used in Testing Programming Accuracy . . . . 34
# LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. FORTRAN Program for Groups with Equal N</td>
<td>25</td>
</tr>
<tr>
<td>2. Sample Data Output for Groups with Equal N</td>
<td>27</td>
</tr>
<tr>
<td>3. FORTRAN Program for Groups with Unequal N</td>
<td>31</td>
</tr>
<tr>
<td>4. Sample Data Output for Groups Having Unequal N</td>
<td>35</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

One of the most perplexing problems facing the novice researcher is that of the statistic to be applied to his collection of data from an experiment. To have an adequate and successful experiment the decision pertaining to the type of statistic to be used must be made "a priori," that is, before starting the experiment. Even though the statistic to be used is of the "a posteriori" class, the decision to use the statistic must be a part of the preliminary planning for the experiment.

Although the examples and treatments that follow are based upon educational and psychological experimentation, the methods are easily adaptable to the physical and social sciences. A change of terminology, from group to plot, test to treatment, and one is able to go from psychological experimentation to biological or chemical experimentation using the same statistic.

Purpose

The primary purpose of this paper is to analyze and evaluate a relatively recently proposed statistic, "Duncan's New Multiple Range Test," for use when the researcher's
plans call for a test of significance between three or more group means. A second is to present computer programs for use with an I.B.M. 1620 digital computer that will accomplish the arithmetic calculations involved in the use of the New Multiple Range Test, and illustrate the use of the Multiple Range Test in testing a research hypothesis. Two programs are proposed, one for groups having equal numbers, the other for groups having unequal numbers in each group.

Historical Background

In the field of research, an assumption sometimes made is that if one does not find a significant difference, the experiment has failed. However, if an experimental design is well planned and the principles are followed, an experiment that results in the discovery of no significant difference may be more important than one in which statistical evidence will verify the researcher's hypothesis. As an oversimplified example, consider the following situation that has been reported by McGuigan (14), McNemar (15), Winer (24), and others.

Suppose that a researcher starts a simple experiment between two groups. After analyzing the data, he finds no affirmation of the hypothesis; therefore, he repeats the experiment with two more groups. This may continue until he succeeds in finding the groups that will give confirmation to his hypothesis, or he may decide to use the data
from one group in one experiment with the data from a group in another experiment. In reporting the results of the experiment, no results are reported on the non-significant groups, or they are not incorporated in the final analysis. In such an instance, the researcher has violated the principles of qualified research by confounding his data, and not taking into account his failures.

The above situation, although not intentional, is applicable to the researcher who decided to use more than two groups and use the Student's $t$ (21) for pairwise comparisons between the means of all the groups. A statistic introduced by Fisher (8), known as the F ratio or simple analysis of variance, will reduce the chance of rejecting a true null hypothesis. It is for this reason that McNemar insists upon the following ultimatum. "When and only when $F$, as an overall test, indicates significant differences among $G$ groups may we safely make further tests to see whether two selected means differ significantly" (14, p. 285).

This stand is taken by a number of statisticians and researchers, but there is an equal, if not greater, number who are drifting away from the above concept. The members of this dissident group are the proponents of the many multiple comparison type of statistics.

Student (21) was probably the first to suggest the use of multiple comparisons. Newman (16) expanded Student's
ideas, and Keuls (12) arranged these ideas into a workable solution based upon the range between extreme means. This is the test widely known as the "Newman-Keuls Studentized Ranges." This is one of the five best known of the multiple comparisons or multiple range tests. The \( l_s d \) (least significant difference), suggested by Fisher (8), is the \( t \) test between all means within the groups after one has found a significant \( F \). Other tests within the multiple comparisons group are: Scheffe's method of judging all contrasts (18), Tukey's \( h_s d \) and \( w \) tests (22), and Duncan's two tests, "Multiple Comparisons" (3) and "New Multiple Range Test" (1).

Multiple comparisons tests are in use in many fields of research, but due to their recent introduction, many researchers refrain from making use of these techniques. This is due in part to the large number of techniques available and the decision as to which test is the most appropriate. A review of special statistical books edited for such fields as agronomy, biology, business, social science, engineering and psychology will show a recommendation of one or more of the above listed multiple comparisons or one-way classification of analysis of variance. In each case, the author proposes the method most appropriate to the particular discipline (6, 7, 9, 11, 14, 17, 20, 24).
Before proceeding to the major problems of this paper, some of the assumptions regarding multiple comparisons should be explained. It should also be noted, this paper is designed to meet the needs of the novice researcher. If a mathematical proof is needed, a reference note will be made concerning the location of the proof.

Multiple comparisons test or one-way classifications of analysis of variance is, in essence, the analysis of the homogeneity of the means of a number of groups or treatments. As an example, suppose a researcher wants to know the effect of a certain method on eight groups with twenty subjects each. One method would be to find the average or mean of each group, then selecting the largest mean to be the best. Another method would be to use the \( t \)-test for pairwise comparisons between each two group means. This procedure known as the "multiple-\( t \)" will lead one into making twenty-eight comparisons. This is not only laborious, but if a five per cent level of significance is set as a criterion, a serious error enters in the results of the experiment. When all comparisons have been made, the probability of rejecting a true hypothesis is no longer five per cent. In making these twenty-eight pairwise comparisons, the probability of finding at least one significant difference is forty per cent. This is the basis of the reasoning behind the ultimatum voiced by McNemar as
quoted earlier. As a protection from making such an error, one should first compute the F ratio, then if significant, perform the t tests.

Wine (23) approaches the problem in a different manner. Since the F test combined with the multiple-t is known as the LSD, when using this method, one protects from one type of error, then returns to the original position of the t tests between all pairs of means. It is for this reason he states the following position.

However, if this principle of using a preliminary F test were carried to its logical conclusion, any test involving a group of means would need to be preceded by a preliminary F test of the group. Obviously, such a procedure becomes unwieldy. Fortunately, nearly the same results can be achieved by replacing the F tests with range tests and doing away with the t test (23, p. 353).

Another factor in favor of the range tests (multiple comparisons) is that at times the F test may show a homogeneity when this is not true. Some very startling examples are presented by Snedecor (19), Federer (7), Steel and Torrie (20), and Wine (23). The following, from Wine (23, p. 354), is an example. Suppose that the extreme means, the lowest and highest, are the same in a three group comparison test; such as, means = 2.0, 2.1 and 3.0, with the variance = 0.41, and F ratio is 3.71. In this case, the null hypothesis would not be rejected. However, in a similar circumstance, with means = 2.0, 2.0 and 3.0, and the same variance, the F ratio becomes 4.17 and is significant.
beyond the 5 per cent level. It seems illogical that one should accept and reject the null hypothesis when the range (2.0 to 3.0) of the means did not change.

Very similar situations may arise when using the multiple comparisons tests. As mentioned previously, the multiple-t introduces too many errors to be considered as a technique to be used with more than three groups. When with three groups one may reach the logical paradox of concluding that mean (one) is not different from mean (two), mean (two) is not different from mean (three), but mean (one) is different from mean (three). Fisher's l s d will provide adequate safeguards for this situation provided the number of groups remain at four or less.

The Newman-Keuls "Studentized Range Test" was the first major attempt to measure significant differences between the means in a ranked situation. The ease with which this statistic may be applied is an asset to its acceptance. The two main objections are: the inability to accommodate groups with an unequal number of subjects in each group, and the criterion used as a test for significance. The Newman-Keuls method uses a comparison criterion that increases as the steps (or groups) between the means compared increases, but it is insensitive to the degrees of freedom for the total number of subjects involved. Its application is that of the t test between means with a larger value to be compared as the steps increase.
Tukey's two tests are an attempt to overcome some of the problems arising from the Newman-Keuls test and multiple-t tests. The first test, the "honest significant difference" (h s d), uses a comparison criterion that is so large that one may fail to find a significant difference between the means when it actually exists. The "within" test relaxes this restriction to a mid-point between the h s d and the Newman-Keuls value for the smallest means, then increases the value as the steps become further apart. In both of Tukey's tests the researcher is restricted to the use of an equal number of subjects in each group. Winer (23) and Wine (22) have both presented long discussions and comparisons of these tests with other methods discussed in this section. Both present tables of comparisons of the critical values showing how each test is adaptable to specified experimental settings.

Scheffe's "S" method requires the F ratio test for overall homogeneity, if this is significant, proceed with comparison of differences between the means. The comparison value is the significant F for the total degrees of freedom and total number of groups involved. With Scheffe's test it is possible to obtain a significant F but no significant difference between the pair of means. It is for this reason many authors recommend using an F at the .10 level for comparison with the difference between the means.
To overcome some of the many problems of multi-group comparisons of mean, Duncan (5) embarked in the field with his thesis. His first effort dealt with the significant differences between means in a ranked order. Duncan (4) next proposed a refinement of his thesis and expanded it to the analysis of variance. The third proposal was the widely known "Multiple Comparisons Test" (3). This particular test is the most powerful technique available to prevent the researcher from making "Type II" errors. The main faults of the Multiple Comparisons Test are its difficult computational procedures and its inability to detect "Type I" errors.

Duncan (1) next proposed the Multiple Range and Multiple F Test which he calls the "New Multiple Range Test." According to Duncan (4, p. 2), it "combines the simplicity and speed of application of the 'Newman-Keuls Test' with most of the power advantages of the multiple comparisons tests." The New Multiple Range Test, as originally proposed, is for groups having an equal number of subjects. Kramer (13) and Duncan (2) have extended the test to handle groups of unequal number of repetitions. Harter (9, 10) found the original "range tables" contained numerous errors, and published corrected and extended tables. (See Appendix.)

Scheffe (18), another worker in the field of multiple comparisons, has this to say of Duncan's method.
I have not included the multiple-comparisons methods of D. B. Duncan because I have been unable to understand their justifications. Duncan was one of the earliest workers in this field. His rules for deciding for all pairs of means whether the means differ (and if so, in which direction) do not have the rapidly increasing insensitivity with increasing \( q \) or \( k \) that may disappoint users of the S- and T-methods, the insensitivity of these being related to the increasing lengths of the intervals for contrasts. The interested reader is referred to an expository paper, Duncan (1955), and subsequent papers, Kramer (1956) and Duncan (1957). The paper, Duncan (1952), giving the justification originally advanced for the methods, contains errors, e.g., the \( F \) in the basic distribution (2) cannot be central \( F \) as claimed (18, p. 78).

The "S- and T-methods" are comparisons tests proposed by Scheffe and Tukey respectively.

Scheffe does not give support to the above statement in any previous or later publications. The above statement is the basis for all other authors who have cited objections to Duncan's methods (11, 15 and 18). It is also the first problem approached by this paper.
CHAPTER BIBLIOGRAPHY


10. __________, "Critical Values for Duncan's New Multiple Range Test," Biometrics, XVI (December, 1960),


CHAPTER II

NEW MULTIPLE RANGE TEST

A statistical procedure must, in itself, be valid before the researcher is able to obtain a valid analysis of his data from an experiment. It is for this reason that the first problem of this paper is to show that Duncan's New Multiple Range Test (2) is a reliable statistic to be used. Scheffe (16), a mathematical statistician, has questioned the methods used by Duncan (4, 5) in deriving his original formulas.

The only section voicing disagreement to the New Multiple Range Test has been quoted in its entirety earlier. Scheffe (16) specifically points out that the "F in the basic distribution cannot be a central F as claimed" (16, p. 78). He was correct in making this assumption. Duncan, in private correspondence, verified the assumption when informed of Scheffe's criticism. This is the reply from Duncan by private correspondence, January 28, 1964.

I did make a mistake in my 1952 paper and quoted a probability density in a central F form when it should have had a non-central F form.

Duncan's 1952 paper (4), by his own admission, contains other errors of an "expository nature," but they were
not of critical importance. When trying to follow the derivations and assuming that a central F form is used, Scheffe was justified in his criticisms. Duncan follows the above statement with an explanation in his correspondence of January 28, 1964.

The computations for the tables involved however were all for significance levels. These are worked out at points in the parameter space where the non-centrality parameter is zero and where the mistake did not affect anything. This is why the mistake did not come to light and also why it has no bearing on the actual test and tables that were published.

In addition to his letter, Duncan sent a copy of his most recent publication (1). This paper, using a Bayes Rule conditional probability function, gives full mathematical justification for the methods proposed in his 1955 and 1957 papers (2, 3).

Scheffe, when informed of these developments, replied that he "should attempt that paper sometime." He went on to explain that he was slow at reading technical publications. As it stands now, McNemar (14, p. 286) does not accept Duncan's method because of the objection voiced by Scheffe. Winer (23, p. 85) remarks that Scheffe "takes issue with the principles underlying the development of the sampling distributions," then he gives the computational methods involved in the New Multiple Range Test and compares it with other multiple comparison tests.
Another point brought out by Scheffe in private correspondence deals with "the probability of reaching various decisions in various situations" (power in two decision cases). Duncan (4, 5) did not develop this point in his earlier papers, but in his 1960 paper (1), this point is well developed. This paper was examined with regard to the points brought out by Scheffe, also in light of some of the problems brought out by Lehmann, a reference used by Duncan. On the basis of Lehmann's general derivations, Duncan was able to make specific applications to the problems of multiple comparisons tests. The two-decision probabilities were calculated by the use of a Bayes Rule for "a priori" decisions.

The specific development by Duncan was a two-decision situation with a non-central $t$ and a non-central parameter. The two-decision case was expanded to a related three-decision case, then to each of $\frac{1}{2} n(n-1)$ pairwise comparisons of an experiment of any size. The 1960 paper (1) gives a rigid mathematical proof of the New Multiple Range Test and the Multiple Comparisons Test; in this paper Duncan (1, p. 1014) proposes to present a new and better test in the future.

When computing the error rates for Duncan's New Multiple Range Test (2), Harter (9) found that the protection level ranges were in error. This error is from using tables
by Pearson and Hartley (15) that contained errors when extended to a multiple error decision. Harter (8) gives the mathematical proof of the original error and computations used in deriving new tables. (See Appendix.) The tables are prepared by the use of Univac Computers, using formulation and programs furnished by Harter and his associates. Many recent publications (6, 7, 12, 13, 18, 20, 22) still have Duncan's (2, 3, 4) original tables. These should be replaced by tables compiled by Harter.

Assumptions

The New Multiple Range Test is a valid technique: the researcher may use this method with no reservations. The only remaining question must be answered on an individual basis with regard to the experiment and the researcher's decision. Should one use a preliminary F test, or should one proceed with the New Multiple Range Test alone?

The assumptions underlying the use of any range or multiple comparisons test are:

(1) the decision to make comparisons between all means,
(2) need of protection from making Type II errors,
(3) testing of non-homogeneous groups,
(4) decisions regarding the "best" of four or more groups.

The new test is based on "protection levels" not significance levels; therefore, the question regarding Type I
errors have been minimized. These "protection levels" are based on the non-central parameter and experiment-wise degrees of freedom. Basically, all multiple comparisons tests are designed so as to furnish a protection from making Type II errors; that is, accepting the null hypothesis when it is false. The protection levels proposed by Duncan are shown in tables in the Appendix, these tables being for significance level $\alpha = .05, .01$ and $0.001$. By examining these tables, it is apparent that as the degrees of freedom increase, the protection value decreases. If the degrees of freedom remain constant and the steps (groups) increase, the protection levels increase. The New Multiple Range Test will not detect Type I errors as readily as the F test, but when the experimental design requires the use of a large number of groups, it provides a reasonable level of significance balanced with the protection levels for Type II errors.

In determining which of the fifteen differences among six treatment means are significant, Steel and Torrie (20) state, "For this procedure, it is not necessary to compute an F value and proceed only if this is significant; the investigator may use the procedure regardless of the significance of F" (20, p. 107). This statement deals with the use of Duncan's test. They continue with the explanation based upon the following assumption: "Since the
notion of Type I errors was not originally intended to apply to multiple comparisons, the idea of significance levels is replaced by that of 'special protection levels' against finding false significant differences" (20, p. 109). This idea is carried over into a later paper, where Steel (19) expands Duncan's method to a sign test and other useful statistical methods.

The position taken by most authors is, if the researcher feels that a preliminary F is necessary, then compute the F. If this is significant, do not revert back to a multiple-t if more than four group means are to be compared. The researcher also should remember that the higher the significant level set for rejecting the null hypothesis, the greater will be the probability of committing Type II errors. The researcher must make the final decision with regard to the circumstances which surround his individual experiment.

Computer Application

I have tried to emphasize the importance of computers in the development of psychology as a rational quantitative science. Within psychology, members of the Psychometric Society generally are well qualified to assume leadership in the introduction of computer activities. Certainly, it is crucial that our graduate students become trained in computer use, and during their training the more intimate their contact with computers the more probably they will display wisdom in adapting computer equipment in future research (10, p. 327).
This introduction by Jones (10), should be a guideline for graduate students in all fields of research. His paper brings to light the many duplicative processes and programs found in computer centers. Computers are not only machines to handle long and tedious calculations, they are also used as means of performing experiments and teaching subjects to perform more adequately.

With the above problem in mind, this paper was designed to acquaint the students of psychology with some of the problems in computer programming. In the process of programming any type of mathematical calculation, one must adhere to the specific type of language and calculating methods of the computer at hand. The FORTRAN (FORmula TRANslator) language is an arithmetic type of statement designed specifically for digital computers. Some of the particular requirements of programming will be explained later in this section, particular attention being called to the fact that some letters used in defining variables to the computer must be used in a specific manner.

One of the first problems found when programming Duncan's New Multiple Range Test was the different formulas for use with groups having equal and unequal numbers of subjects in each group. The most important factor in this instance was time. Operating time of the computer is of greater importance than long and complicated calculations.
Since most experiments are carried out with groups having an equal number of subjects in each group, a program was designed for this specific purpose. With this program, the researcher who has his data for a twenty group design could get the results of the statistical analysis approximately two minutes after entering the data in the computer. If the groups were of an unequal number in each group, the time factor would be about five times as long.

Specifically, Duncan's New Multiple Range Test was programmed in FORTRAN for an I.B.M. 1620, 20k digital computer. The programs presented here are designed to accommodate up to a twenty-group set of data. This is an arbitrary limit and may be changed if necessary. By changing the "dimension statement," these programs may be extended to the limiting number of groups set by the "critical value" tables of one hundred groups. (See Appendix.) This limiting factor does not apply to the number of subjects within the separate groups. The total number within each group is decided by the researcher, not the computer.

Program Explanation

Mention was made previously of the requirement that certain letters have specific meanings in computer programming. When using FORTRAN, the letters J, K, L, M, N specify digits used in the limiting process of the computer. These letters are used with "whole numbers" and cannot become a
part of the calculation except by making special instruction to the computer. For this reason, the mathematical symbols are used to instruct the computer. The use of M to indicate mean, n to indicate number within the group, and G to indicate group is not compatible with the FORTRAN language. In using letters and numbers to define variables in FORTRAN, it is necessary to use all capital letters. The symbols used in these programs conform to standard FORTRAN and as close to the statistical symbols as is possible. The following definition of symbols gives the mathematical notation, program notation and appropriate statistical term.

Definition of symbols:

\[ G = M \quad \text{number of groups} \]
\[ n = N \quad \text{number within each group} \]
\[ \sum x = XS = \text{sum of } x \]
\[ \sum x^2 = X2 = \text{sum of } x^2 \]
\[ SS = SS = \text{sum of squares} \]
\[ Se = SE = \text{error variance} \]
\[ r_p = RT = \text{Range table (critical range values)} \]
\[ R_p = RP = \text{Range product} \]

The last two symbols, \( r_p \) and \( R_p \), are the designations from Duncan's papers (2, 3, 4) and used by other authors (6, 7, 14, 18).

The following is an explanation of the statements used in the program shown in Figure 1. (See page 25.)
The "dimension statement" sets up storage location for the derived variables and determines the maximum size of the program. Statements numbered 50 to 55 instruct as to output form.

Statement 1 indicates the number of groups.

Statements between 1 and 3 set up and perform looping operations to calculate the sum of X (XS) and sum of X squared (X2).

Statement 3 computes the mean (XB) for each group.

Statement 2 computes the sum of squares (SS), and prepares the data for computing the error variance (SE).

Statement 4 inputs the range table values (RT).

Statement 20 tests the equality of N between each of the groups. If the N's are not equal, the computer types "USE ALTERNATE PROGRAM - FOR EQUAL N," then stops.

If the N's are equal, the computer continues with:

Statement 21 computes the range product (RP).

Statements 22 through 27 sorts the means (XB) into ranks of ascending order.

Statements 6, 7, and 8 output instructions for data.

Data Preparation

There are four types of data to be introduced into the computer. These are: the number of groups (M), the number within each group (N), the variates (X), such as, scores or trials and the significance level values from the tables of critical values (RT), called range table values. The data are to be prepared in the following manner.
1st card - - - Number of groups (M)
2nd card - - - Number in first group (N)
3rd card - - - Variates (X)

For each group to follow there must be one N card followed by card, or cards, with the variates (X) of the group. The last card for input is to be the range table (RT) card. The data for this card are found in the table of Critical Values. (See Appendix.) Since these values are based on the "degrees of freedom," they are found in the following manner:

\[ df = T - M \]

where T is the total number of subjects (sum of N's) and M is the number of groups. For example, suppose the research design has 6 groups with 21 subjects in each group, with a significance level \( \alpha = .05 \).

\[ df = 126 - 6 = 120 \]

Enter the table in the \( \nu \)-column at \( \nu = 120 \). These values are to be used on the RT card. (Note: there is no \( p = 1 \).) The values for \( p = 2,6 \) would be entered on the RT card as follows:

\[
0.000 \quad 2.800 \quad 2.947 \quad 3.045 \quad 3.116 \quad 3.172
\]

The RT card must have the complete row associated with the df (degrees of freedom) for as many means as are to be compared. Tables for significance level \( \alpha = .05, .01 \) and \( .001 \).
are presented in the Appendix. These tables are so in nature that it is possible to make linear interpolation between the given values.

Since most experiments are carried out with an equal number in each group, this is the main program. If all groups are not equal, the computer will type "USE ALTERNATE PROGRAM - FOR UNEQUAL N," then stop. This is a protective device in case one of the X variates is lost from a data card. The computations for groups having an unequal number in each group differ from that of equal N.

Figure 1 gives the complete computer program for equal N. The numbers preceding each statement give the location in the computer for the arithmetic statement which follows. Immediately following the program is a sample of the output of results, shown in Figure 2, using data from McGuigan (13, pp. 178-180) to test the accuracy of the programming.
Fig. 1--A FORTRAN program for groups with equal N

```
C
C COCO PIG
C
C DUNCANS MULTIPLE RANGE TEST
C FOR GROUPS WITH EQUAL N

DIMENSION XB(20), RT(20), XN(20), RP(20), T(20)

50 FORMAT (/10X 26HDUNCANS RANGE TEST RESULTS///)
51 FORMAT (5X37HUSE ALTERNATE PROGRAM - FOR UNEQUAL N/)
52 FORMAT (10 F8.3/)
53 FORMAT (10X 4HMEAN/)
54 FORMAT (/10 X11HRANKED MEAN//)
55 FORMAT (/10X14H RANGE PRODUCT//)

1 READ, M
A=0.
B=0.
DO 2 J=1,M
XS=0.
X2=0.
READ , N
DO 3 I=1,N
READ, X
XS= XS+X
3 X2 = X2+(X**2)
XN(J)=N
SS=X2-(XS*XS)/XN(J)
XB(J)=XS/XN(J)
A= A+SS
2 B=B+(XN(J)-1.)
```
SE = SQRTF (A/B)

DO 4 J=1,M

4 READ, RT(J)

DO 5 J=2,M

IF (XN(J) - XN(J-1)) 20, 21, 20

20 TYPE 51

STOP

21 RP(J) = SE*RT(J)*SQRTF(1./XN(J))

5 CONTINUE

TYPE 50

TYPE 53

DO 6 J=1,M

6 TYPE 52, XB(J)

DO 24 I=1,M

T(I) = XB(I)

J=1

IF (I-1) 27, 24, 22

22 IF (T(J)-T(J-1)) 23, 24, 24

D=T(J-1)

T(J-1)=T(J)

T(J)=D

IF (J-2) 24 , 24, 25

J=J-1

GO TO 22

CONTINUE

TYPE 54

DO 7 I=1,M

7 TYPE 52, T(I)
The sample output from the computer, shown in Figure 2 is arranged in the following manner. The first row, under (MEAN), shows the mean of each group as it was entered into the computer. Next, the Ranked Mean, shows the means of the groups, ranked in ascending order. The Range Product was calculated and output is for comparing the differences between groups.

Fig. 2—Sample data output for groups with equal N

DUNCANS RANGE TEST RESULTS

<table>
<thead>
<tr>
<th>MEAN</th>
<th>23.900</th>
<th>7.200</th>
<th>1.100</th>
<th>23.500</th>
<th>1.500</th>
</tr>
</thead>
<tbody>
<tr>
<td>RANKED MEAN</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1.100</td>
<td>1.500</td>
<td>7.200</td>
<td>23.500</td>
<td>23.900</td>
</tr>
<tr>
<td>RANGE PRODUCT</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.219</td>
<td>1.283</td>
<td>1.326</td>
<td>1.352</td>
<td></td>
</tr>
</tbody>
</table>
To make comparisons of the differences between the ranked means and range product values, start with the highest and lowest means, groups 5 and 1, and a significance level $\alpha = .05$. This compares the extreme means first, if and only if, this difference is greater than the range product for five groups. Test the remaining groups, each time using the range product associated with the steps between the groups being compared. The following procedure is recommended:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Means</th>
<th>Range Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 - 1</td>
<td>23.90 - 1.100 = 23.80</td>
<td>$&gt; 1.352$</td>
</tr>
<tr>
<td>5 - 2</td>
<td>23.90 - 1.500 = 22.40</td>
<td>$&gt; 1.326$</td>
</tr>
<tr>
<td>5 - 3</td>
<td>23.90 - 7.200 = 16.20</td>
<td>$&gt; 1.283$</td>
</tr>
<tr>
<td>5 - 4</td>
<td>23.90 - 23.500 = 0.40</td>
<td>$&lt; 1.219^{**}$</td>
</tr>
<tr>
<td>4 - 1</td>
<td>23.50 - 1.100 = 22.40</td>
<td>$&gt; 1.326$</td>
</tr>
<tr>
<td>4 - 2</td>
<td>23.50 - 1.500 = 22.00</td>
<td>$&gt; 1.283$</td>
</tr>
<tr>
<td>4 - 3</td>
<td>23.50 - 7.200 = 16.30</td>
<td>$&gt; 1.219$</td>
</tr>
<tr>
<td>3 - 1</td>
<td>7.20 - 1.100 = 6.10</td>
<td>$&gt; 1.283$</td>
</tr>
<tr>
<td>3 - 2</td>
<td>7.20 - 1.500 = 5.70</td>
<td>$&gt; 1.219$</td>
</tr>
<tr>
<td>2 - 1</td>
<td>1.50 - 1.100 = 0.40</td>
<td>$&lt; 1.219^{**}$</td>
</tr>
</tbody>
</table>

(**not significant)

These results show that in comparing all means, groups four and five are not significant, and groups one and two do not differ significantly. In all other comparisons, there is a significant difference between the means of the groups. A method of showing this is to underline the means of those groups that are not significantly different. This is shown in Figure 2 where means of groups four and
five and groups one and two are underlined to show there is no significant difference between these means.

Alternate Program

Most of the operations in the alternate program are the same as in the regular program. There are two specific differences. The first involves the ranking of means in ascending order. This must be done by the computer operator. The computer is capable of ranking the means, but it is not possible to keep the N in the same locations corresponding to the proper mean. The Mean and N must be used together, as will be shown later.

The procedure to follow is to first input M, N, and X as in the regular program. The computer will compute the mean of each group, type out the means, then type out the following statements:

"RANK GROUPS WITH MEANS IN ASCENDING ORDER" "RESUBMIT DATA"

The computer operator will then rearrange the groups so that the means will start with the lowest and go to the highest, making sure that the N card remains with the proper group data. Resubmit the data (M, N, X). The RT card is added as the last card at this time. Do not enter the RT card before this time.

The computation will be the same until computing the range product (RP). This is the main difference between
the two programs. In the regular program (for equal N), the range product (RP) is the product of the error variance, range table value and the square root of the reciprocal of N.

\[ RP = SE \times RT \times \sqrt{1/N} \]

For the alternate program, use the square root of 
\[ \sqrt{\frac{1}{N_a} + \frac{1}{N_b}} \]

\[ RP = SE \times RT \times \sqrt{\frac{1}{N_a} + \frac{1}{N_b}} \]

where a and b are the two groups being compared. Since the computer is not able to randomly choose groups for comparison, the procedure is to perform the calculation between all groups in sequence with the means ranked in ascending order. This is the reason for ranking the groups before computing the data.

If it is assumed the protection level of significance level \( \alpha = .05 \), then the results are shown in Figure 4. (See page 35.) This follows the FORTRAN program for unequal N, shown in Figure 3. The alternate program is programmed using the same variable as is used in the regular program. This listing is made with the output typewriter of the computer in the same manner as the regular program. Although this program is designed to accommodate up to twenty groups, as in the regular program, it may be extended to handle up to one hundred groups.
Fig. 3--FORTRAN program for groups with unequal N

```
C DUNCANS MULTIPLE RANGE TEST
C FOR GROUPS WITH UNEQUAL N
C ALTERNATE PROGRAM

DIMENSION XB(20), RT(20), XN(20), RP(20)
50 FORMAT(/10X 26HDUNCANS RANGE TEST RESULTS///)
51 FORMAT (/)
52 FORMAT (6F14.4/)  
53 FORMAT (10X 4HMEAN///)
54 FORMAT (/10 X11HRANKED MEAN///)
55 FORMAT (/10X14H RANGE PRODUCT///)
56 FORMAT (/5X20H RANK GROUPS WITH MEANS IN ASCENDING ORDER/)  
57 FORMAT (/5X20H RESUBMIT DATA CARDS/) 
101 READ , MM
102 DO 103 J= 1,MM 
103 XS=0.
104 READ, N
105 DO 106 I=1,N
106 READ, X
107 XS = XS +X
108 XN(J) = N
109 XB(J) = XS/XN(J)
110 TYPE 53
111 DO 112 J=1,MM
112 TYPE 52, XB(J)
113 TYPE 56
```
09182 TYPE 57
09206 PAUSE
09218 1 READ, M
09242 A=0.
09266 B=0.
09280 DO 2 J=1,M
09302 XS=0.
09326 X2=0.
09350 READ, N
09374 DO 3 I=1,N
09386 READ, X
09410 XS= XS+X
09446 3 X2 = X2+(X*X2)
09530 XN(J)=N
09578 SS=X2-(XS*XS)/XN(J)
09674 XB(J)=XS/XN(J)
09758 A= A+SS
09794 2 B=B+(XN(J)-1.)
09902 SE = SQRTF (A/B)
09950 DO 4 J=1,M
10046 4 READ, RT(J)
10046 TYPE 50
10070 TYPE 55
10094 DO 7 J=2,M
10106 N=J-1
10142 DO 5 K=1,N
10154 L=J-K+1
As a method of testing the accuracy of the above program and a means of illustrating the use of the New Multiple Range Test in testing a research hypothesis the scores presented in Table I which appears on the following page were used. These are scores made by applicants for employment at a large industrial firm in Fort Worth, Texas. This particular set of scores is from an overall validation study of the effectiveness of the applicant test in use to predict job success. The purpose of this part of the overall research was to make pairwise comparisons between the mean scores made by a set proportion of the employees in six randomly selected departments.

In order that a random design be obtained the following procedure was used to determine the scores to be used. Each department was assigned a two digit number, beginning with 00, and each employee within the department was assigned a three digit number beginning with 000. Selection of the department and members within selected departments
was made by the use of a table of random numbers. Since the range of education of the employees was so wide, from seventh grade to Doctor of Philosophy, the range of scores was large. The New Multiple Range Test was used since the set proportion would give an unequal number of scores in each of the groups.

TABLE I
DATA USED IN TESTING PROGRAMMING ACCURACY

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 26</td>
<td>9 19</td>
<td>21 32</td>
<td>50 53</td>
<td>70 53</td>
<td>35 47</td>
</tr>
<tr>
<td>28 20</td>
<td>25 25</td>
<td>18 24</td>
<td>51 46</td>
<td>71 66</td>
<td>65 68</td>
</tr>
<tr>
<td>12 18</td>
<td>15 23</td>
<td>24 22</td>
<td>43 41</td>
<td>63 61</td>
<td>89 68</td>
</tr>
<tr>
<td>22 21</td>
<td>31 27</td>
<td>23 24</td>
<td>42 52</td>
<td>62 72</td>
<td>56 82</td>
</tr>
<tr>
<td>16 17</td>
<td>23 17</td>
<td>24 23</td>
<td>34 52</td>
<td>77 54</td>
<td>77 41</td>
</tr>
<tr>
<td>20 19</td>
<td>25 29</td>
<td>33 31</td>
<td>34 48</td>
<td>72 54</td>
<td>82 62</td>
</tr>
<tr>
<td>18 15</td>
<td>13 11</td>
<td>23 23</td>
<td>52 53</td>
<td>68 72</td>
<td>77 82</td>
</tr>
<tr>
<td>9 16</td>
<td>27 27</td>
<td>24 27</td>
<td>40 33</td>
<td>53</td>
<td>56 87</td>
</tr>
<tr>
<td>33 17</td>
<td>25 25</td>
<td>14 31</td>
<td>41 32</td>
<td>71 67</td>
<td>83 87</td>
</tr>
</tbody>
</table>

The results of this evaluation are shown in Figure 4. It is readily apparent that this set of results presents more output from the computer than the results from groups having equal N. (See Figure 2, page 27.) The arrangement of results in Figure 4 are as follows: first, group means in order of input into the computer followed by directions to the operator to sort the data and resubmit it to the
Fig. 4—Sample data output for groups having unequal N

The data listed under Range Product are for comparing the means between groups. The first row is for comparing mean two with one; the second row, mean three with one, three with two; the third row, mean four with one, four with two, four with three; this continues until all means have been compared for significant differences. There is one exception to this process—test the extreme means first, that is, the largest and smallest. If this difference is not significant, then no further test may be made. Duncan states, "If and only if the difference between the extreme means is significant, then may further tests be made" (1, p.2).
Analysis of Data

The results from the Range Test (Figure 4) are analyzed in the following manner. Starting with the extreme means, make pairwise comparisons between all means taking into account the number of steps between the means:

<table>
<thead>
<tr>
<th>Groups</th>
<th>Means</th>
<th>Range Product</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 - 1</td>
<td>69.1000 - 19.6470</td>
<td>49.4530 &gt; 6.6766</td>
</tr>
<tr>
<td>6 - 2</td>
<td>69.1000 - 21.5555</td>
<td>47.4445 &gt; 6.4595</td>
</tr>
<tr>
<td>6 - 3</td>
<td>69.1000 - 24.4736</td>
<td>44.6264 &gt; 6.2243</td>
</tr>
<tr>
<td>6 - 4</td>
<td>69.1000 - 43.2083</td>
<td>25.8917 &gt; 5.6931</td>
</tr>
<tr>
<td>6 - 5</td>
<td>69.1000 - 64.5333</td>
<td>4.5667 &lt; 6.1023**</td>
</tr>
<tr>
<td>5 - 1</td>
<td>64.5333 - 19.6470</td>
<td>44.8863 &gt; 7.0431</td>
</tr>
<tr>
<td>5 - 2</td>
<td>64.5333 - 21.5555</td>
<td>42.9778 &gt; 6.7924</td>
</tr>
<tr>
<td>5 - 3</td>
<td>64.5333 - 24.4736</td>
<td>40.0597 &gt; 6.4947</td>
</tr>
<tr>
<td>5 - 4</td>
<td>64.5333 - 43.2083</td>
<td>21.3250 &gt; 5.8803</td>
</tr>
<tr>
<td>4 - 1</td>
<td>43.2083 - 19.6470</td>
<td>23.5613 &gt; 6.1590</td>
</tr>
<tr>
<td>4 - 2</td>
<td>43.2083 - 21.5555</td>
<td>21.6528 &gt; 5.8631</td>
</tr>
<tr>
<td>4 - 3</td>
<td>43.2083 - 24.4736</td>
<td>18.7346 &gt; 5.4862</td>
</tr>
<tr>
<td>3 - 1</td>
<td>24.4736 - 19.6470</td>
<td>5.8266 &lt; 6.2776**</td>
</tr>
<tr>
<td>3 - 2</td>
<td>24.4736 - 21.5555</td>
<td>2.9181 &lt; 5.8764**</td>
</tr>
<tr>
<td>2 - 1</td>
<td>21.5555 - 19.6470</td>
<td>1.9085 &lt; 6.0422**</td>
</tr>
</tbody>
</table>

(** not significant)

These results imply that one would fail to reject the null hypothesis when comparing group six to five, and between groups one, two and three. The results are shown in Figure 4 (See page 35) by underlining the non-significant group means. Any two groups not joined by the same line are significantly different from each other. In this case, one would infer that groups five and six received better scores...
than all other groups at a level of significance of better than the .05 level. Groups one, two and three were not significantly different from each other, and group four was significantly different from all groups.

Although the procedure outlined here may appear to be confusing, one should take note of the fact that the comparisons start with the highest and lowest means. The Range Product output is arranged so that the RP needed is directly above the smaller mean of the two groups being compared. The last (fifth) row in the Range Product output is used for comparing the highest mean with all other group means. In this case group six is compared first. The next step is to compare group five mean with the lesser means, using the values in the fourth row and so on until all comparisons have been made.

It is possible to have the computer calculate the differences between the means. It was felt that the addition of these differences at this time would be confusing to the experimenter. When Duncan's method becomes more widely accepted and used, then the program may be expanded to do the differencing. For the present, the researcher should perform this operation until he is thoroughly acquainted with the use of the Multiple Range Test.

Computation Example

In some instances the researcher may have need of the computational formulas used in Duncan's New Multiple Range
Test; therefore, this section is presented to show how the formulas are utilized and to check the computer program. The data used here will be the same as presented in Table I. Using the same method of maintaining a randomized design as presented earlier, this validation study is to make pairwise comparisons between the mean scores obtained on an applicant’s test by selected personnel in selected departments of an industry. This is a part of an overall study designed to set up standards for selection to the various departments, replace or supplement the present test, and if necessary, omit the testing of certain types of applicants for employment. Since this particular part of the overall study was only a validation analysis of previously completed parts of the study, no conclusions were drawn except that the results presented in Figure 4 were similar to that found in the other parts of the study.

For this example, suppose that the research calls for a pairwise comparison between six groups to (1) determine the ability of the applicant test to predict job success of the employee, (2) if the test is not discriminatory enough, exempt certain of the applicants from the testing situation, and (3) determine if the present test should be supplemented with another type of test.

So as to retain continuity, the symbols used here will be the same as in the computer programs. The first
step is to compute the sum of squares (SS) for each group. To do this, first compute the sum of X (XS), sum of X squared (X2), and the mean (XB) for each group. The sum of squares for any group is found by

\[ SS = X^2 - \frac{(XS)^2}{N} \]

For each of the groups this is:

\[
SS_1 = 7110 - \frac{(334)^2}{17} = 547.9 \\
SS_2 = 9082 - \frac{(388)^2}{18} = 718.5 \\
SS_3 = 11781 - \frac{(465)^2}{19} = 400.7 \\
SS_4 = 46219 - \frac{(1037)^2}{24} = 1412.0 \\
SS_5 = 63386 - \frac{(968)^2}{15} = 917.7 \\
SS_e = 100212 - \frac{(1382)^2}{20} = 4715.8 \\
Total SS_1 = 8712.6
\]

Add the sum of squares for all groups to get the numerator of the error variance, this shown as the total above. The error variance (SE) is found by:

\[
SE = \sqrt{\frac{\sum SS_1}{\sum (N_i - 1)}} = \sqrt{\frac{8712.6}{107}} = 9.024
\]

Next, compute the degrees of freedom (df) as shown previously,

\[ df = 113 - 6 = 107 \]
Enter the table of critical values (significance level = .05) shown in the Appendix; since 107 is close enough to 120, interpolation was not used. The RT values for this example are:

\[
RT = 2.800 \ 2.947 \ 3.045 \ 3.116 \ 3.172
\]

If the groups had contained an equal number of scores, the computing of RP would have been greatly simplified. In this example, the number of scores in each group differs; therefore to calculate RP the following formula is used,

\[
RP = SE \times RT \times \sqrt{\frac{1}{N_a} \frac{1}{N_b}}
\]

where \(N_a\) and \(N_b\) are the two groups being compared, and the RT is the value for the number of steps between the groups.

Compare the group with highest mean with each of the other groups starting with the least mean. The means in a ranked order are:

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.65</td>
<td>21.55</td>
<td>24.47</td>
<td>43.21</td>
<td>64.53</td>
<td>69.10</td>
</tr>
</tbody>
</table>

\[
\begin{align*}
RP_{e,1} &= (9.02)(3.172) \sqrt{\frac{1}{2}(1/20+1/17)} = 6.6694 \\
RP_{e,2} &= (9.02)(3.116) \sqrt{\frac{1}{2}(1/20+1/18)} = 6.4673 \\
RP_{e,3} &= (9.02)(3.045) \sqrt{\frac{1}{2}(1/20+1/19)} = 6.2111 \\
RP_{e,4} &= (9.02)(2.947) \sqrt{\frac{1}{2}(1/20+1/24)} = 5.6911 \\
RP_{e,5} &= (9.02)(2.800) \sqrt{\frac{1}{2}(1/20+1/15)} = 6.0894** \\
RP_{5,1} &= (9.02)(3.116) \sqrt{\frac{1}{2}(1/15+1/17)} = 7.0578 \\
RP_{5,2} &= (9.02)(3.045) \sqrt{\frac{1}{2}(1/15+1/18)} = 6.7871 \\
RP_{5,3} &= (9.02)(2.947) \sqrt{\frac{1}{2}(1/15+1/19)} = 6.4889 \\
RP_{5,4} &= (9.02)(2.800) \sqrt{\frac{1}{2}(1/15+1/24)} = 5.8620 \\
RP_{4,1} &= (9.02)(3.045) \sqrt{\frac{1}{2}(1/24+1/17)} = 6.1551
\end{align*}
\]
\[ \text{RP}_{4,2} = (9.02)(2.947) \sqrt{\frac{1}{2}(1/24 + 1/18)} = 5.8506 \]
\[ \text{RP}_{4,3} = (9.02)(2.800) \sqrt{\frac{1}{2}(1/24 + 1/19)} = 5.4830 \]
\[ \text{RP}_{3,1} = (9.02)(2.947) \sqrt{\frac{1}{2}(1/19 + 1/17)} = 6.2761^{**} \]
\[ \text{RP}_{3,2} = (9.02)(2.800) \sqrt{\frac{1}{2}(1/19 + 1/18)} = 5.8620^{**} \]
\[ \text{RP}_{2,1} = (9.02)(2.800) \sqrt{\frac{1}{2}(1/18 + 1/17)} = 6.0388^{**} \]

(\text{**not significant})

By arranging these results in the same form as the computer output in Figure 4, a comparison of these results may be made with computer results.

\begin{center}
\begin{tabular}{c c c c}
\hline
Range & Product \\
\hline
6.0388 & 6.2761 & 5.8620 & 5.4830 \\
6.1551 & 6.8506 & 5.4820 & 5.8620 \\
7.0578 & 6.7871 & 6.4889 & 5.8620 \\
\hline
\end{tabular}
\end{center}

Although these results differ slightly from that shown in Figure 4, the difference (caused by rounding off during calculating) is not large enough to affect the results obtained when comparing the difference between the means and the range product. By using the same procedure as outlined when analyzing the results of Figure 4, it was found that group six does not differ significantly from group five, and that groups one, two and three do not differ significantly from each other. These results would seem reasonable since all the members of groups five and six were college graduates, employed in professional positions. Members of groups one, two and three had an average education level of ten years, while group four had an average of thirteen years of school.
CHAPTER BIBLIOGRAPHY


CHAPTER III

SUMMARY

The purpose of this paper is to analyze and evaluate a statistical technique, Duncan's New Multiple Range Test, for use when the research design calls for pairwise comparisons between all means in a multi-group experiment. A second purpose was to propose and present computer programs designed for use in a digital computer, and a third was to illustrate the use of Duncan's New Multiple Range Test in testing a research hypothesis.

The validity of Duncan's method has been questioned by Scheffe, a mathematical statistician. Upon the basis of Scheffe's criticism, other authors have voiced disapproval for Duncan's method. Since very few publications are available concerning this controversy, it was necessary to conduct private correspondence with the principal authors, Duncan and Scheffe. The two letters pertinent to the solution of this problem are reproduced as a part of the Appendix. Duncan's latest paper (1) gives a mathematical proof of his method by use of the Bayes Rule for conditional probabilities. This paper, along with previous publications (2, 4), interject a new approach to the use of the multiple-\( t \) test after obtaining a significant \( F \) ratio. The researcher who uses
the multiple-\textit{t} must face the hazards of rejecting a true null hypotheses or accepting false research hypotheses. The probability of accepting false research hypotheses mounts rapidly with the number of groups compared, when the experiment contains more than three or four groups. Duncan's New Multiple Range Test will offer an adequate protection for this situation. Some of the more prominent multiple comparisons tests are noted in the historical background and references are presented for the interested reader.

The original tables proposed by Duncan (2, 4) were found to contain errors. These errors were found by Harter (5, 6), while calculating error rates in a probability function, and new tables were published. A copy of the three most useful tables, significance level \( \alpha = .05, .01 \) and \( .001 \), are reproduced in the Appendix.

When programming Duncan's New Multiple Range Test for electronic computers, it was necessary to use separate programs for groups having equal and unequal numbers within the groups. Since the majority of experiments are carried out with an equal number in each group, the main program was designed for this purpose. The programs are designed to do all mathematical calculations except the final differencing between the means of the groups. The researcher has only to subtract one mean from another, compare this result with the derived "least significant range" and if the difference
is larger, there is a significant difference between the groups compared.

The program for groups with unequal numbers in each group, labeled "Alternate Program," makes use of the extensions by Kramer (7) and Duncan (3). The main difference between the two methods is the formula for finding the least significant range, labeled "range product." There is a separate range product derived for each comparison between any two groups. The significant difference between the means are determined according to this difference, being either greater than or less than the derived range product. Another important variation is that the computer operator must arrange the groups so that the means of the groups are in ascending order. The computer derives the mean for each group, types out the mean for all groups, then pauses until the groups are sorted by the computer operator. All data are then resubmitted for calculation. As in the regular program, the differencing is computed by the researcher.

During the process of programming the New Multiple Range Test, a number of peculiarities were noted regarding the limitations of the computer. Although there are two programs proposed, this does not imply that groups with equal N's cannot be used in the Alternate Program. In fact, the Alternate Program being used in all cases causes
difficulties to arise from the fact that when all group N's are equal there is a large amount of duplication of output. The duplication of numbers in the Range Product would lead to excessive confusion when comparing with the differences between the means.

A computational example, for groups with an unequal number is presented as an illustration of testing a research hypothesis. This example uses the same data as was used to test programming accuracy; therefore, an accuracy check for the program is available. The computation follows the outline presented by McGuigan (8, pp. 172-187). This outline was also used in programming the two computer programs.
CHAPTER BIBLIOGRAPHY


APPENDIX

1. Critical Value Tables
2. Private Correspondence (Duncan)
3. Private Correspondence (Scheffe)
Critical Values for Duncan's New Multiple Range Test

<table>
<thead>
<tr>
<th>r</th>
<th>p</th>
<th>Protection Level P = (0.05)**</th>
<th>Significance Level α = 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

|   |   | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 2 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 3 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 4 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 5 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 6 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 7 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 8 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 9 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

**Critical Values for Duncan's New Multiple Range Test**

**Protection Level** \( P = (0.05)^*; Significance Level \( a = 0.05 \)

---

**Protection Level** \( P = (0.01)^*; Significance Level \( a = 0.01 \)

---

### Critical Values for Duncan's New Multiple Range Test

**Protection Level $P = 0.05$**

**Significance Level $\alpha = 0.01$**

<table>
<thead>
<tr>
<th>$r$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003 0.003</td>
</tr>
<tr>
<td>7</td>
<td>0.049 0.049 0.049 0.049 0.049 0.049 0.049 0.049 0.049 0.049 0.049 0.049 0.049 0.049 0.049 0.049 0.049 0.049 0.049 0.049</td>
</tr>
<tr>
<td>8</td>
<td>0.079 0.079 0.079 0.079 0.079 0.079 0.079 0.079 0.079 0.079 0.079 0.079 0.079 0.079 0.079 0.079 0.079 0.079 0.079 0.079</td>
</tr>
<tr>
<td>9</td>
<td>0.095 0.095 0.095 0.095 0.095 0.095 0.095 0.095 0.095 0.095 0.095 0.095 0.095 0.095 0.095 0.095 0.095 0.095 0.095 0.095</td>
</tr>
<tr>
<td>10</td>
<td>0.109 0.109 0.109 0.109 0.109 0.109 0.109 0.109 0.109 0.109 0.109 0.109 0.109 0.109 0.109 0.109 0.109 0.109 0.109 0.109</td>
</tr>
<tr>
<td>11</td>
<td>0.121 0.121 0.121 0.121 0.121 0.121 0.121 0.121 0.121 0.121 0.121 0.121 0.121 0.121 0.121 0.121 0.121 0.121 0.121 0.121</td>
</tr>
<tr>
<td>12</td>
<td>0.131 0.131 0.131 0.131 0.131 0.131 0.131 0.131 0.131 0.131 0.131 0.131 0.131 0.131 0.131 0.131 0.131 0.131 0.131 0.131</td>
</tr>
</tbody>
</table>

Mr. Walter Wheeler  
916 Willowwood Drive  
Denton, Texas  

Dear Mr. Wheeler:  

In reply to your letter of late January, I did make a mistake in my 1952 paper and quoted a probability density in a central F form when it should have had a non-central F form. The computations for the tables involved however were all for significance levels. These are worked out at points in the parameter space where the non-centrality parameter is zero and where the mistake did not affect anything. This is why the mistake did not come to light and also why it has no bearing on the actual test and tables that were published.

In the same paper there were indeed other mistakes of an expository nature, but none again having any bearing on the critical issues involved.

The main reference I think you will need is my Biometrics paper "Multiple range and multiple F tests" Vol. 11 1955. Better tables for this have been put out by Dr. Leon Harter ref. (4) on the enclosed list. My 1955 paper is the one giving the test to which I think McGuigan and Winer refer.

(I am enclosing a more recent paper in this area which in a sense replaces the exposition in the 1952 paper to which Scheffe refers.)

I trust this is of some help.

Sincerely,

David B. Duncan

DBD/hb
Dear Mr. Wheeler:

Thank you for your recent letter and the bibliography. I have not read the 1961 Annals paper, but the title suggests it requires the assumption of a priori probabilities for application, and thus I do not usually find practicable. However, I should attempt that paper again soon. I did look at the 1955 Biometrics paper and found nothing much that I was looking for; I want to know the operating or performance characteristic (power in two decision case), that is, the probability of reaching various decisions in various situations; few probabilities seem to be calculated there. Perhaps my reluctance to tackle Doornik's papers can be
understood by my explanation that I am a painfully slow reader of technical material, and I struggled a long time with his early papers, with negative results.

Perhaps Professor cit's view, who is mentioned in the Annual's paper has refuted the claims of the method. Should you find any such authority in mathematical statistics I would be grateful for your informing me.

Thanks again,

Schefe

P.S. May I congratulate you on your attitude about trying to verify somehow the validity of proposed statistical tools in your profession! — HS
BIBLIOGRAPHY

Books


Articles


Unpublished Materials


Duncan, David B., Private Correspondence, 1963.

Scheffe, Henry, Private Correspondence, 1963.