J.D. WILLIAMS

OCTOBER 1976

VOLUME 7 NUMBER 1

MULTIPLE LINEAR REGRESSION VIEWPOINTS

A publication of the Special Interest Group on Multiple Linear Regression

MONOGRAPH SERIES #2

MULTIPLE COMPARISONS BY MULTIPLE LINEAR REGRESSIONS

MULTIPLE LINEAR REGRESSION VIEWPOINTS

Multiple Linear Regression Viewpoints
Vol. 7, No. 1, 1976
Monograph Series #2

MULTIPLE COMPARISONS BY MULTIPLE LINEAR REGRESSION

John D. Williams

The University of North Dakota

Multiple Linear Regression Viewpoints
Vol. 7, No. 1, 1976
Monograph Series #2

Multiple Linear Regression Viewpoints
Vol. 7, No. 1, 1976
Monograph Series #2

ABSTRACT

Several of the more common multiple comparison techniques are explored in a regression approach. Dunnett's test for comparing several groups to a single group, Tukey's(a) honestly significant different test, Newman-Keul's, Tukey's(b) and Duncan's tests are considered. Complex comparisons (contrasts) are shown through Dunn's and Scheffé's tests and through orthogonal comparisons. Orthogonal polynomials are also shown for testing for trend. A method for finding a maximized Scheffé contrast such that the contrast will yield the same R^2 value as the original full model is also included.

The intent of the present monograph is to more fully explore the use of alternate methodologies to the usual multiple F tests when more than one restriction is placed on a full model.

Multiple Linear Regression Viewpoints
Vol. 7, No. 1, 1976
Monograph Series #2

TABLE OF CONTENTS

Multiple Linear Regression Viewpoints
Vol. 7, No. 1, 1976
Monograph Series #2

CHAPTER I

INTRODUCTION: DUNNETT'S TEST

Several researchers have presented multiple linear regression as a data analysis technique. A major impetus to this approach was the publication by Bottenberg and Ward (1963). Jennings (1967) continued in the same vein in an article concerning the two-way fixed effects analysis of variance from a regression viewpoint. Ward (1969) compared four different approaches to data analysis and showed that all four approaches had many basic ideas in common. The difficulty of recognition of the relationship between the use of regression analysis and standard analysis of variance designs was pointed out by Cohen (1968). A more recent statement on their approach to regression was made by Ward and Jennings (1973).

Similar approaches to the Bottenberg-Ward-Jennings have also been made. Kelly, Beggs and McNeil (1969) have elaborated this approach with several educational research applications; this particular presentation has recently been updated (McNeil, Kelly and McNeil, 1975). The text by Williams (1974a) also has its origin in the Bottenberg-Ward-Jennings approach.

Other authors have somewhat similar book length treatments on regression as well. Mendenhall (1969), Kerlinger and Pedhazur (1974) and Cohen and Cohen (1975) are prominent examples of such tests.

Multiple Comparisons

The rejection of the null hypothesis $\mu_1 = \mu_2 = \mu_3 . . . = \mu$ would rarely seem to be completely satisfying. It is only natural to ask, if there are differences, where are those differences? Since the overall

F test does not provide the solution, then it would seem natural to seek a direct solution to the researcher's problem(s). If multiple use of the t test were made, the reported probability level would be seriously violated; hence the use of multiple t tests would be inappropriate. What is approprate depends upon the situation. Several multiple comparison tests have been devised to satisfy a particular demand or analysis upon the data. Thus, it would seem that there is no best multiple comparison method that universally fits all situations; rather, if the researcher can point out the comparisons of interest, then an appropriate multiple comparison method can be chosen.

The present monograph is oriented toward examining several of the more often used multiple comparison procedures. Thus, Dunnett's (1955, 1964) test for several treatment groups with a control, Tukey's (1953), Dunn's (1961), Duncan's (1955) Scheffé's (1953), Newman (1939) and Keuls' (1952) tests and the orthogonal contrasts are separately considered.

Error Rates in Regression

One consideration within the regression framework that still needs additional concern is the consideration of error rates. Commonly a full model is specified and then one or more restricted models are tested. If more than one restricted model is specified, then an appropriate adjustment should be made concerning the probability level. To be specific about the adjustment of the probability levels, it is helpful to consider differentiating among five terms relating to multiple comparisons: error rate per comparison, error rate per experiment, experimentwise error rate, error rate per family and the familywise error rate.

These error rates have been defined as follows:

$$\text{Per comparison} = \frac{\text{No. of comparisons incorrectly called significant}}{\text{total number of comparisons}}\,; \qquad (1.1)$$

$$\text{Per experiment} = \frac{\text{No. of comparisons incorrectly called significant}}{\text{total number of experiments}}\,; \qquad (1.2)$$

$$\text{Experimentwise} = \frac{\text{No. of experiments containing erroneous statements of significance}}{\text{total number of experiments}}\,; \qquad (1.3)$$

$$\text{Per family} = \frac{\text{No. of comparisons incorrectly called significant}}{\text{total number of statements in the family}}\,; \qquad (1.4)$$

$$\text{Familywise} = \frac{\text{No. of experiments containing erroneous statements of significance}}{\text{total number of statements in the family}}\,. \qquad (1.5)$$

It can be seen that in general the first three error rates will be different, with the error rate per comparison being the least stringent, and the error rate per experiment being the most stringent. Ryan (1959, 1962) has indicated that the use of the per comparison error rate should be discouraged, and that one of the other approaches should be adopted. An example discriminating among the first three types of error rates follows; discussion of the family and familywise error rates is postponed until after the first three are discussed.

Suppose a set of six comparisons of four means are contemplated: $\overline{X}_1$ to $\overline{X}_2$, $\overline{X}_1$ to $\overline{X}_3$, $\overline{X}_1$ to $\overline{X}_4$, $\overline{X}_2$ to $\overline{X}_3$, $\overline{X}_2$ to $\overline{X}_4$ and $\overline{X}_3$ to $\overline{X}_4$. Also 100 replications are made of this experiment. Suppose also in the <u>population</u> all hypotheses are true. For the sample data, suppose the following results are found:

Experiments with		Number of Incorrect Rejections
zero incorrect rejections	89	0
one incorrect rejection	4	4
two incorrect rejections	3	6
three incorrect rejections	0	0
four incorrect rejections	1	4
five incorrect rejections	2	10
six incorrect rejections	1	6
	100	30

Then, the per comparison error rate is $\frac{30}{600} = .05$; the experimentwise error rate is $\frac{11}{100} = .11$; and the per experiment error rate is $\frac{30}{100} = .30$.

In general, the per experiment error rate is the most conservative in that a higher critical value is usually required than the other two error rates. Also, experimenters are encouraged to use the experiment as the unit of analysis; while the experimentwise error rate or per experiment error rate are seen as being acceptable, the per comparison error rate is viewed as being generally unacceptable (Ryan, 1962). The per comparison error rate is very similar to using multiple t tests.

The Family Error Rates

For a full discussion of family error rates, Miller (1966) can be consulted. In particular, Miller points out the ambiguity of the term "family". In actual practice, if the family of comparisons of interest are identical to the comparisons made in the experiment, then the familywise error rate is identical to the experimentwise error rate and the per family error rate is identical to the per experiment error rate. If, however, more than one family is defined for an experiment, this relationship no longer holds.

For example, if a two-way analysis of variance is performed, the experimenter may define three families of comparisons: one family for rows, one family for columns and one family for interactions. Within a family of comparisons (e.g., rows) an experimentwise .05 level might be maintained. If the entire experiment is considered as the unit of analysis, and if each family is tested at the .05 level, then the experimentwise error rate for the complete experiment is 3(.05) = .15.

A Priori and A Posteriori Tests

Another consideration in regard to multiple comparison tests is the concept of a priori and a posteriori tests. An a priori test is one in which the comparisons have been decided on in advance of the gathering of data. On the other hand, a posteriori tests allow the researcher to decide on the hypotheses to be tested even after the data has been inspected. Thus, the a posteriori tests allow the researcher the flexibility to consider relationships of interest after a preliminary data analysis has been completed. The cost of this flexibility is loss in power; that is, the tests where the researcher has set the hypotheses to be tested in advance of the data collection will generally have somewhat more power and will allow the null hypothesis to be rejected more often than if the hypotheses have not been set in advance.

A very crude, but useful rule to follow in choosing between an a priori test and an a posteriori test is the following: if you know what you're doing, use an a priori test; otherwise, use an a posteriori test. In other words, the a priori tests are quite useful in theory testing and theory building situations; the a posteriori tests are more of a general purpose type of test.

One and Two Tailed Tests

A natural extension of using a priori tests is the consideration of using a one-tailed test. McNeil and Beggs (1971) previously have considered the use of directional hypotheses in multiple linear regression. Some multiple comparison methods lend themselves to one-tailed tests when the comparisons have been posited a priori. If the experimenter is willing to predict the direction of an outcome of a posited test on an a priori basis, then a more powerful test can result. The reader should be cautioned that, with the exception of Dunnett's test, all tables in the Appendix to this monograph are two-tailed tests. The tables can be used for one-tailed tests by halving the reported probabilities.

Dunnett's Test for Comparisons of Several Treatment Groups with a Control

Dunnett (1955, 1964) devised a test that would allow the comparisons of several treatment groups with a control group and still retain an experimentwise error rate. This test could also be used whenever an experimenter wished to test a group which might be called the "experimental group" against several existing (but different) groups.

For example, a business educator may have devised a new approach to teaching beginning typewriting. The business educator may find that instead of finding one typical approach to teaching typewriting there may be several methods being used. Rather than lumping all of the existing methods together and calling them a control group, it would seem more logical to test the new approach against each existing group separately, but in a single experiment. Dunnett's test is appropriate for this situation. So that the various tests can be compared to one another, a single data set is used throughout this monograph. That data set is given in Table 1.1.

TABLE 1.1

DATA FOR DUNNETT'S TEST

Control Group Group One	Group Two	Group Three	Group Four
9	8	13	15
8	7	10	12
6	8	12	10
3	6	11	17
4	6	14	11

$\overline{X}_1 = 6.0$, $\overline{X}_2 = 7.0$, $\overline{X}_3 = 12.0$, $\overline{X}_4 = 13.0$.

Suppose the interest is in comparing the Control Group to Groups Two, Three and Four.

Viewing the problem from a regression viewpoint, it is helpful to define four binary predictors:

X_1 = 1 if the score is from a member of the control group (Group One): and 0 otherwise,

X_2 = 1 if the score is from a member of Group Two; and 0 otherwise,

X_3 = 1 if the score is from a member of Group Three; and 0 otherwise, and

X_4 = 1 if the score is from a member of Group Four; and 0 otherwise.

A linear model can be written for this situation:

$$Y = b_0 + b_2X_2 + b_3X_3 + b_4X_4 + e_1, \tag{1.6}$$

where

b_0 = the Y-intercept,

b_2 = the regression coefficient for Group Two,

b_3 = the regression coefficient for Group Three,

b_4 = the regression coefficient for Group Four, and

e_1 = the error involved in prediction.

It can be noticed that the control group has seemingly been left out. However, if equation 1.6 is solved for the expected value for a member of the control group,

$$E(Y) = b_0 + b_2(0) + b_3(0) + b_4(0),$$

$$E(Y) = b_0.$$

The expectancy for a member of the control group is by definition $\overline{X}_1$. Thus, a least squares solution for b_0 is $\overline{X}_1$, the mean of the control group.

For a member in Group Two, the expected value is

$$E(Y) = b_0 + b_2(1) + b_3(0) + b_4(0),$$

$$E(Y) = b_0 + b_2,$$

$$E(Y) = \overline{X}_1 + b_2. \tag{1.7}$$

A least squares solution for the expectancy of a given member of Group Two is the mean of Group Two. Thus

$\overline{X}_2 = \overline{X}_1 + b_2$, from equation 1.7, or

$$\overline{X}_2 - \overline{X}_1 = b_2. \tag{1.8}$$

Likewise

$$b_3 = \overline{X}_3 - \overline{X}_1 \text{ and } b_4 = \overline{X}_4 - \overline{X}_1.$$

Equation 1.6 can be rewritten

$$Y = \overline{X}_1 + (\overline{X}_2 - \overline{X}_1)X_2 + (\overline{X}_3 - \overline{X}_1)X_3 + (\overline{X}_4 - \overline{X}_1)X_4 + e_1. \tag{1.9}$$

Equation 1.9 lists precisely the comparisons of interest for comparing several treatments with a control. Since equation 1.6 (and, therefore, equation 1.9) is the same model as has been given for a one-way analysis of variance (Williams, 1971, 1974a), this approach also yields results identical to the analysis of variance situation. Thus, using equation 1.6, it can be seen that these two useful results can be obtained simultaneously: the usual analysis of variance as one part of the output, and Dunnett's test as the other part.

The information necessary for a regression solution, with equation 1.6 as the linear model, can be conveniently placed in tabular form (see Table 1.2).

TABLE 1.2

REGRESSION FORMULATION FOR COMPARING SEVERAL TREATMENTS WITH A CONTROL

Y	X_1	X_2	X_3	X_4
9	1	0	0	0
8	1	0	0	0
6	1	0	0	0
3	1	0	0	0
4	1	0	0	0
8	0	1	0	0
7	0	1	0	0
8	0	1	0	0
6	0	1	0	0
6	0	1	0	0
13	0	0	1	0
10	0	0	1	0
12	0	0	1	0
11	0	0	1	0
14	0	0	1	0
15	0	0	0	1
12	0	0	0	1
10	0	0	0	1
17	0	0	0	1
11	0	0	0	1

TABLE 1.3

OUTPUT OF MULTIPLE REGRESSION PROGRAM

Variable No.	Mean	Standard Deviation	Correlation X vs Y	Regression Coefficient	Standard Error of Regression Coefficient	Computed t Value	Beta
3	0.25000	0.44426	-0.40109	1.00001	1.36014	0.73522	0.12033
4	0.25000	0.44426	0.40109	6.00001	1.36014	4.41130	0.72197
5	0.25000	0.44426	0.56153	7.00001	1.36014	5.14652	0.84230
Dependent 1	9.50000	3.69210					
Intercept		5.99999					
Multiple Correlation		0.84515					
St. Error of Estimate		2.15058					

Analysis of Variance for the Regression

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F Value
Attributable to Regression	3	185.00027	61.66675	13.33340
Deviation from Regression	16	73.99973	4.62498	
Total	19	259.00000		

For the data in Table 1.2, a general purpose multiple regression program was used. Table 1.3 contains the printout from that analysis. The variable number in Table 1.3 refers to the <u>order</u> in Table 1.2; the criterion variable is variable number 1; variable 2 refers to the control group, variable 3 to Group Two, variable 4 to Group Three and variable 5 to Group Four. Because variable 2 refers to the Control Group, no information appears in the printout using that variable number. The table of residuals has not been included herein.

Table 1.3 contains the previously mentioned items. It can be recalled that $\overline{X}_1 = 6.0$, $\overline{X}_2 = 7.0$, $\overline{X}_3 = 12.0$, $\overline{X}_4 = 13.0$. The intercept is 6.0 (within rounding error) and is $\overline{X}_1$. Also, $b_2 = 1 = \overline{X}_2 - \overline{X}_1$, and is in keeping with equation 1.9. Similar statements could be made concerning b_3 and b_4. The computed t values in Table 1.3 are identically the same values as would result from the use of Dunnett's test. It is only necessary to compare each of these values to Dunnett's table for the test of significance. From Table Id, a computed t value of 3.39 is needed for significance at the .01 level on a two-tailed test. Thus, both Groups Three and Four are significantly higher than the Control Group. It thus can be seen that the computed t values, which are tests of the partial regression weights, should be evaluated in this instance <u>not</u> by the traditional t table, but by use of Dunnett's tables.

CHAPTER II

MAKING ALL SIMPLE COMPARISONS: TUKEY'S, NEWMAN-KEULS' AND DUNCAN'S TESTS

This chapter considers four tests that differ only in their use of probability; from a regression viewpoint, the hypotheses being tested are identical. The four tests are Tukey's (a) test, sometimes called the honestly significant different test, Newman-Kuels' test, Tukey's (b) test and Duncan's multiple range test. The tests are normally used if all $\binom{k}{2}$ simple comparisons are of interest; the tests are often employed on an *a posteriori* basis. For purposes of comparison among these four tests and also for comparing these tests to the other tests, the data given in Table 1.1 are used.

Tukey's (a) Test: The Honestly Significant Different (HSD) Test

A useful test which can be used on an *a posteriori* basis and retain an experimentwise error rate is Tukey's (1953) HSD test. Quite often, Tukey's test is used when a significant F value has been found, and the interest is in pinpointing where the differences are. It is not necessary that the overall F test be applied to use Tukey's test, however. Also, Tukey's test is usually applied to all pairs $\binom{k}{2}$ of means in the situation where there are more than two groups. Thus, if there were four groups, then there would be $\binom{4}{2}$ or 6 comparisons.

For the data presented in Table 1.1, there are $\binom{4}{2}$, or 6 comparisons of interest (that is, all possible comparisons of pairs) for Tukey's test.

They are the following:

$\overline{X}_1$ to $\overline{X}_2$,

$\overline{X}_1$ to $\overline{X}_3$,

$\overline{X}_1$ to $\overline{X}_4$,

$\overline{X}_2$ to $\overline{X}_3$,

$\overline{X}_2$ to $\overline{X}_4$ and

$\overline{X}_3$ to $\overline{X}_4$.

There are several ways that Tukey's test can be achieved, even using a regression approach. The particular method described here has one major advantage over other descriptions of the calculations of Tukey's test: it is by far the easiest to accomplish, using a computer. A much more elegant solution is given by Williams (1972). The present solution was first given in Williams (1974b).

A Simplified Solution to Tukey's Test

Tukey's test can be accomplished with k - 1 successive uses of regression equations (where k is the number of groups) where models similar to equation 1.6 are solved; in each successive solution, a different group is omitted from the equation.

Thus, for the data in Table 1.1, the solution given for Dunnett's test also includes three of the six comparisons of interest for Tukey's test. Each of the succeeding models also include three of the six comparisons of interest.

To be more specific, Group Two can be omitted from the model, yielding

$$Y = b_0 + b_1X_1 + b_3X_3 + b_4X_4 + e_1. \quad (2.1)$$

Similarly, Group Three can be omitted from the model, yielding

$$Y = b_0 + b_1X_1 + b_2X_2 + b_4X_4 + e_1. \quad (2.2)$$

It should be noted that the values for b_0 and the remaining regression coefficients will be different for equations 1.6, 2.1 and 2.2 The value for b_0 is equal to the mean of the of the group "left out" of the model.

For equation 2.1, $b_0 = \overline{X}_2 = 7.0$; for equation 2.2, $b_0 = \overline{X}_3 = 12.0$.
As indicated, b_1, b_2, b_3 and b_4 will in general be different from equation to equation. As an example, b_4 occurs in equations 1.6, 2.1 and 2.2. For equation 1.6, $b_4 = \overline{X}_4 - \overline{X}_1 = 13 - 6 = 7$; for equation 2.1, $b_4 = \overline{X}_4 - \overline{X}_2 = 6$; for equation 2.2, $b_4 = \overline{X}_4 - \overline{X}_3 = 1$. As is indicated here, the regression coefficients are the difference between the means of each group to the group left out in a particular analysis. However, the value for e_1 will remain the same for each of the equations, as each of these equations can be conceptualized as a reparameterization of an analysis of variance model.
The computed t values for equation 2.1 can be found for the following comparisons:

$t = -.735$ for comparing X_1 to X_2 (this was also available from the first model, equation 1.6);

$t = 3.676$ for comparing $\overline{X}_3$ to $\overline{X}_2$; and

$t = 4.411$ for comparing $\overline{X}_4$ to $\overline{X}_2$.

The computed t values for equation 2.2 can be found for the following comparisons:

$t = -4.411$ for comparing $\overline{X}_1$ to $\overline{X}_3$ (this was also available from the first model, equation 1.6);

$t = -3.676$ for comparing $\overline{X}_2$ to $\overline{X}_3$ (this was also available from the second model, equation 2.1); and

$t = .735$ for comparing $\overline{X}_4$ to $\overline{X}_3$.

A change in sign from one equation to the next in computed t values is to be expected. In the first model, whose results are shown in Table 1.3, the computed t value for comparing $\overline{X}_1$ to $\overline{X}_2$ is .735. From equation 2.1, the computed t value for comparing $\overline{X}_2$ to $\overline{X}_1$ is -.735. This is because the means are being compared in a different order. The first test shows $\overline{X}_2 > \overline{X}_1$. As we would surely expect, the second test showed $\overline{X}_1 < \overline{X}_2$.

To evaluate these t values, the Tables of the Studentized range statistic are to be used, but a modification is first necessary. Using such a table, the value of q for 4 groups and $df_w = 16$, at the .01 level is $q = 5.192$. To evaluate the computed t values, the value for q should be divided by $\sqrt{2}$: the critical value $= \frac{q}{\sqrt{2}} = \frac{5.192}{1.414} = 3.671$. A modified Studentized Range table is included as Tables IIa and IIb in the Appendix. These tables have been modified so that the computed t values can be tested directly; they have been constructed so that the values are calculated from $\frac{q}{\sqrt{2}}$ where q was the value from the original Studentized range statistic tables. The values in Tables IIa and IIb can be used to construct confidence limits in a manner analogous to ordinary confidence limits using the usual t table. Table 2.1 contains the results for Tukey's test (omitting minus signs).

TABLE 2.1

TUKEY'S TEST IN A REGRESSION FORMULATION

Comparison	t
$\overline{X}_1$ to $\overline{X}_2$	.735
$\overline{X}_1$ to $\overline{X}_3$	4.411*
$\overline{X}_1$ to $\overline{X}_4$	5.147*
$\overline{X}_2$ to $\overline{X}_3$	3.676*
$\overline{X}_2$ to $\overline{X}_4$	4.411*
$\overline{X}_3$ to $\overline{X}_4$	.735

*Significant at the .01 level

Four of the six comparisons are significant at the .01 level.

A computer program has been written (Lindem and Williams, 1975) that automatically calculates Tukey's test for all possible simple comparisons. Results are given in both t and q values, so that conventional tables can be used if the researcher so desires.

The Newman-Keuls' Test

The Newman-Keuls' test is one of several multiple range tests (others considered in ths chapter are Tukey's (b) test and Duncan's test). The Newman-Keuls' test controls the error rate experimentwise only for the complete null hypothesis. While the nominal level is α, the actual experimentwise rate can rise to $\frac{k\alpha}{2}$ where k is the number of groups.

The same linear models are necessary with the Newman-Keuls' test as was used in the Tukey (a) test. To complete the Newman-Keuls' test, the means should first be ordered from lowest to highest: as the original means were $\overline{X}_1 = 6$, $\overline{X}_2 = 7$, $\overline{X}_3 = 12$ and $\overline{X}_4 = 13$, the order from lowest to highest means is $\overline{X}_1$, $\overline{X}_2$, $\overline{X}_3$ and $\overline{X}_4$. A t table can be constructed using the computed t values. See Table 2.2.

TABLE 2.2

COMPUTED t VALUES FOR DATA IN TABLE 1.1
FOR NEWMAN-KEULS' TEST

	$\overline{X}_1$	$\overline{X}_2$	$\overline{X}_3$	$\overline{X}_4$
$\overline{X}_1$		.735	4.411*	5.147*
$\overline{X}_2$			3.676*	4.411*
$\overline{X}_3$				.734

*Significant at the .01 level

First, the largest mean ($\overline{X}_4$) is compared to the smallest mean ($\overline{X}_1$); the computed t value, 5.147 > 3.671, the t_4 value at the .01 level with v = 16, r = 4 from Table IIb. Hence, this difference is significant at the .01 level. Had this difference been non-significant, no further tests would be considered; if the lowest mean is not significantly less than the highest mean, then Newman-Keuls reasoned that no other means should be different from one another either (assuming equal sample sizes).

In that the first test is significant, means that are in a range that include three means can be tested. That is, $\overline{X}_1$ is compared to $\overline{X}_3$ and $\overline{X}_2$ is compared to $\overline{X}_4$. In this case, the critical t value is found from Table IIb with v = 16 and r = 3; the critical t_3 is 3.384. The computed t value for comparing $\overline{X}_1$ to $\overline{X}_3$ is 4.411 ($p < .01$); also, the same t value (4.411) is found for comparing $\overline{X}_2$ to $\overline{X}_4$ ($p < .01$). Because both tests are significant, all tests that include adjacent means can now be entertained. Had no significances been found at this stage, the testing would have been concluded. If only one of the tests showed significance (say $\overline{X}_1$ to $\overline{X}_3$), then only those comparisons in the same row as the comparison judged significant would be of interest; in that case, the remaining comparisons would have been $\overline{X}_1$ compared to $\overline{X}_2$ and $\overline{X}_2$ compared to $\overline{X}_3$.

In that both of the previous comparisons were significant, all adjacent means are of interest for testing. The critical t value for adjacent means is given by v = 16, r = 2; t_2 = 2.921. Only the comparison of $\overline{X}_2$ to $\overline{X}_3$ is significant at the .01 level; t = 3.676.

In reviewing the process, the t_4 for Newman-Keuls' test, 3.671, is identical to the critical value for Tukey's (a) test. However, t_3 = 3.384 is smaller than the value for Tukey's (a) test; t_2 = 2.921 is in fact identical to the critical t value for the usual t test. The superior power of the Newman-Keuls' test is due to using a less stringent error rate than an experimentwise level for all comparisons undertaken.

Tukey's (b) Test - Tukey's Compromise

Tukey's (a) test given earlier was based upon building simultaneous confidence limits for each group. If the interest is in completing significance tests, then the order of the groups can be used, much as was done in the Newman-Keuls' test. Tukey's (b) test differs in only one detail from the Newman-Keuls' test--instead of using the values from the Studentized range as has Newman-Keuls', the Tukey (b) test uses the midpoint of the corresponding values for the Tukey (a) HSD test and the Newman-Keuls' test.

Table 2.3 contains the required t values for significance at the .01 level for Tukey's (a) HSD test, the Newman-Keuls' test and Tukey's (b) test.

TABLE 2.3

VALUES OF THE t REQUIRED FOR SIGNIFICANCE
AT THE .01 LEVEL FOR TUKEY'S (a) HSD TEST,
NEWMAN-KEULS' TEST AND TUKEY'S (b) test

r	2	3	4
Tukey's (a) HSD	3.671	3.671	3.671
Newman-Keuls'	2.921	3.384	3.671
Tukey's (b)	3.296	3.528	3.671

The critical value for Tukey's (b) test can be seen to be the mean of the corresponding critical values for the Tukey (a) HSD test and the Newman-Kuels' test. In applying the Tukey (b) test, the testing procedure is identical to Newman-Keuls' "layer" procedure. Under these restrictions, the test maintains an experimentwise error rate, according to Ryan (1959).

Duncan's Multiple Range Test

Perhaps one of the more controversial approaches to multiple comparisons has been due to Duncan. The controversy revolves around his use of the so-called "protection levels." To understand the use of protection levels, it is somewhat useful to attempt to reconstruct the thinking that might have been involved in arriving at Duncan's multiple range test. Two different researchers may independently test hypotheses using a t with $\alpha = .05$. If we incorporate the two findings into a single interpretation, the probability of a Type I error is $1 - (.95)^2$ or .0975. Duncan would argue that the appropriate error rate for a nominal .05 level would be .0975 when two tests have been performed. The general form of the protection level with a nominal .05 error rate is $1 - (.95)^{k-1}$ where k is the number of groups and the comparisons are made in the layer fashion of Newman-Keuls.

If several values are solved for the .05 nominal level (or .05 protection level, as Duncan would term it), the experimentwise error rates would be

for two means, $1 - (.95)^{2-1} = .05$;
for three means, $1 - (.95)^{3-1} = .0975$;
for four means, $1 - (.95)^{4-1} = .1426$;
for five means, $1 - (.95)^{5-1} = .1855$;
for six means, $1 - (.95)^{6-1} = .2262$ and
for seven means, $1 - (.95)^{7-1} = .2649$.

Duncan then used the Studentized range statistic by finding the point that corresponded: to .05 for $r = 2$, to .0975 for $r = 3$; to .1426 for $r = 4$, . . . to .2649 for $r = 7$.

The use of protection levels has already been adequately criticized in the statistical literature and there is no need to continue that criticism here. The interested reader could consult Miller (1966), Scheffé (1959) and Ryan (1959).

The regression solution for Duncan's test is identical to that of Tukey's test and the Newman-Keuls' test. The only difference in the method occurs when the tables are consulted. Table IIIa and IIIb record necessary t values for significance with Duncan's test at the .05 and .01 levels respectively.

The t values given earlier in Table 2.2 could also be evaluated by Duncan's test. Table 2.4 contains the information given earlier in Table 2.3 and also contains the critical values for Duncan's test.

TABLE 2.4

VALUES OF t REQUIRED FOR SIGNIFICANCE
AT THE .01 LEVEL FOR TUKEY'S (a) HSD TEST,
NEWMAN-KEULS' TEST, TUKEY'S (b) TEST AND DUNCAN'S TEST

r	2	3	4
Tukey's (a) HSD	3.671	3.671	3.671
Newman-Keuls'	2.921	3.384	3.671
Tukey's (b)	3.296	3.528	3.671
Duncan's	2.921	3.047	3.129

While Duncan's test is clearly the more powerful, it achieves this power through the use of protection levels rather than through some other mechanism.

Comparing the Four Tests

The four tests (Tukey's (a), (b), Newman-Keuls', Duncan's) have one thing in common: all have a more stringent error rate than the per comparison error rate. However, only Tukey's tests retain an experimentwise error rate for the total set of comparisons. Also, only Tukey's (a) HSD test will allow the construction of confidence intervals. All four tests also utilize the Studentized range statistic in defining the probability levels. The way in which the Studentized range statistic is used is different for each of the four tests, however. The comparisons involved in the present chapter were such that all groups had equal frequencies. If unequal frequencies occur, strictly speaking, none of the tests are appropriate as the tests assume equal sample sizes. However, Tukey's (a) HSD would seem to be robust under the violation of unequal sample sizes. If all simple comparisons are of interest (and no complex comparisons are of interest) and the frequencies in each group are not too unequal (admittedly a value judgment), the Tukey's (a) HSD becomes an approximate test. If the cell frequencies are quite different, then perhaps Dunn's test or Scheffé's test (see Chapter Four) would be more appropriate.

CHAPTER III

ORTHOGONAL CONTRASTS

A multiple comparison method that seems to have a fairly wide acceptance is the use of orthogonal contrasts. Descriptions of the process are given in several widely used texts (for example, Hays, 1973, Winer, 1971 and Edwards, 1972). The popularity of the orthogonal methodology is due to the fact that contrasts constructed within this framework are independent of one another. Actually, orthogonality is of significant use beyond its use in multiple comparisons; when the cell frequencies in the two-way analysis of variance are proportional, the two main effects and the interaction are independent (or orthogonal); also, orthogonal polynomials are useful in trend analysis.

Constructing Orthogonal Contrasts

A contrast (i.e., comparison) is defined by

$d_i = a_1\overline{X}_1 + a_2\overline{X}_2 + . . . + a_k\overline{X}_k$ where $\sum_1^k a_i = 0$.

For the data in Table 1.2, if $a_1 = 1$, $a_2 = -1$, $a_3 = 0$ and $a_4 = 0$, the resulting equation is $d_1 = \overline{X}_1 - \overline{X}_2$ and is a contrast. Consider another set of values for the a_is: Let $a_1 = 0$, $a_2 = 0$, $a_3 = 1$ and $a_4 = -1$. Not only is the resulting equation, $d_2 = \overline{X}_3 - \overline{X}_4$, a contrast, but d_1 is orthogonal to d_2. To see this, it is helpful to put the contrast coefficients in a table. See Table 3.1.

TABLE 3.1

ORTHOGONAL COEFFICIENTS FOR TWO CONTRASTS

	d_1	d_2
a_1	1	0
a_2	-1	0
a_3	0	1
a_4	0	-1

The contrasts d_1 and d_2 are orthogonal when the products of the corresponding d_i coefficients sum to zero; for the data in Table 3.1, the sum is $(1)(0) + (-1)(0) + (0)(1) + (0)(-1) = 0$.

Finding orthogonal coefficients is no easy task, particularly for those who are not mathematically inclined. A maximum of k-1 such contrasts can be found when there are k groups. A third contrast can be made using $a_1 = ½$, $a_2 = ½$, $a_3 = -½$ and $a_4 = -½$. To show d_3 is orthogonal to both d_1 and d_2, the corresponding values for the a_i coefficients have to be multiplied and then summed. To show d_1 is orthogonal to d_3,

$$(1)(½) + (-1)(½) + (0)(-½) + (0)(-½) = 0.$$

Also, d_2 is orthogonal to d_3:

$$(0)(½) + (0)(½) + (1)(-½) + (-1)(-½) = 0.$$

Interestingly, d_3 is the <u>only</u> contrast that is orthogonal to both d_1 and d_2. On the other hand, several contrasts can be found that are orthogonal to d_1 or d_2 (but not orthogonal to both).

Suppose a second set of orthogonal contrasts is sought. For d_1, let $a_1 = -½$, $a_2 = ½$, $a_3 = -½$ and $a_4 = ½$.

For d_2, let $a_1 = -½$, $a_2 = ½$, $a_3 = ½$ and $a_4 = -½$. It can be seen that

d_1 is orthogonal to d_2:

$$(-\tfrac{1}{2})(-\tfrac{1}{2}) + (\tfrac{1}{2})(\tfrac{1}{2}) + (-\tfrac{1}{2})(\tfrac{1}{2}) + (\tfrac{1}{2})(-\tfrac{1}{2}) = 0.$$

As mentioned earlier, only one d_i will be orthogonal to both d_1 and d_2. It is easily shown that d_3 for this set is identical to the d_3 from the first set: $a_1 = \tfrac{1}{2}$, $a_2 = \tfrac{1}{2}$, $a_3 = -\tfrac{1}{2}$ and $a_4 = -\tfrac{1}{2}$. Table 3.2 contains both sets of orthogonal contrasts.

TABLE 3.2

TWO SETS OF ORTHOGONAL COEFFICIENTS WITH FOUR GROUPS

	Set 1			Set 2		
	d_1	d_2	d_3	d_1	d_2	d_3
a_1	1	0	$\tfrac{1}{2}$	$-\tfrac{1}{2}$	$-\tfrac{1}{2}$	$\tfrac{1}{2}$
a_2	-1	0	$\tfrac{1}{2}$	$\tfrac{1}{2}$	$\tfrac{1}{2}$	$\tfrac{1}{2}$
a_3	0	1	$-\tfrac{1}{2}$	$-\tfrac{1}{2}$	$\tfrac{1}{2}$	$-\tfrac{1}{2}$
a_4	0	-1	$-\tfrac{1}{2}$	$\tfrac{1}{2}$	$-\tfrac{1}{2}$	$-\tfrac{1}{2}$

Set 1 can be used to show the regression solution for orthogonal contrasts. New variables X_5, X_6 and X_7 can be constructed to correspond to d_1, d_2 and d_3: X_5 = 1 if from a member in Group One, - 1 if from a member in Group Two, 0 otherwise;

X_6 = 1 if from a member in Group Three, -1 if from a member in Group Four, 0 otherwise; and

X_7 = $\tfrac{1}{2}$ if from either Groups One or Two, $-\tfrac{1}{2}$ if from either Groups Three or Four. The complete formulation is given in Table 3.3.

TABLE 3.3

REGRESSION FORMULATION FOR ORTHOGONAL CONTRASTS

Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
9	1	0	0	0	1	0	½
8	1	0	0	0	1	0	½
6	1	0	0	0	1	0	½
3	1	0	0	0	1	0	½
4	1	0	0	0	1	0	½
8	0	1	0	0	-1	0	½
7	0	1	0	0	-1	0	½
8	0	1	0	0	-1	0	½
6	0	1	0	0	-1	0	½
6	0	1	0	0	-1	0	½
13	0	0	1	0	0	1	-½
10	0	0	1	0	0	1	-½
12	0	0	1	0	0	1	-½
11	0	0	1	0	0	1	-½
14	0	0	1	0	0	1	-½
15	0	0	0	1	0	-1	-½
12	0	0	0	1	0	-1	-½
10	0	0	0	1	0	-1	-½
17	0	0	0	1	0	-1	-½
11	0	0	0	1	0	-1	-½

The results of using X_5, X_6 and X_7 as predictors of Y yields $R = .84515$, $R^2 = .71429$. The computed t values are $t_1 = -.735$ (for d_1); $t_2 = .735$ (for d_2) and $t_3 = -6.238$ (for d_3). These t values are traditionally compared to an ordinary t table (this practice bears additional comment in the next section): only t_3 is significant ($p < .01$). Also $r_{Y5} = r_{Y6}$ .09825, $r^2_{Y5} = r^2_{Y6} = .00965$ and $r_{Y7} = .83666$, $r^2_{Y7} = .69499$. Then

$$r^2_{Y5} + r^2_{Y6} + r^2_{Y7} = .00965 + .00965 + .69499 = .71429 = R^2.$$

Further, $R^2(SS_T) = .71429\ (259.00) = 185.00$; thus $R^2(SS_T)$ can be seen to be identical to the value for SS_A if a simple analysis of variance is performed. Any complete set of k - 1 orthogonal contrasts will have the property of having $R^2(SS_T) = SS_A$.

Comments on the Use of Orthogonal Contrasts

Three areas need further comment regarding the use of orthogonal contrasts. First, the error rate, experimentwise, is $1 - (1 - \alpha)^{k-1}$. This error rate is identical to the error rate for Duncan's "protection level." Indeed, Duncan's best defense for the concept of the "protection level" is that it was in agreement with the same probabilities associated with orthogonal contrasts. On the other hand, each contrast is independent of all other contrasts, so that each finding might be interpreted independently of all other findings. A second concern is that orthogonal contrasts would seldom seem to be in direct relation to the hypotheses a researcher wished to test. The contrasts of interest generally would not be orthogonal. A third concern is that researchers who are not mathematically oriented may have exceeding difficulty in finding a useful set of orthogonal contrasts. It might appear that the second and third concerns would eliminate orthogonal contrasts as being basic to a researcher's background. Interestingly, the application of orthogonal polynomials to trend analysis is at least one application that satisfies both these concerns. Before explaining orthogonal polynomials, it is useful to consider simple polynomials.

Polynomial Regression

Suppose the data in Table 1.2 (and also in Table 3.3) represent four groups of subjects who have received varying amounts of instruction, and that the Y values represent test scores. Also, Group One subjects have received one unit of instruction; Group Two subjects have received two units of instruction; Group Three subjects have received three units of instruction and Group Four subjects have received four units of instruction. Several new questions could be asked of the data. What type of trend exists? Does a simple linear trend exist, or does a higher ordered polynomial give a better explanation of the data?

Polynomial trends can be given by the equation

$$Y = b_0 + b_1X + b_2X^2 + b_3X^3 + . . . b_{k-1}X^{k-1} + e_2. \tag{3.1}$$

Here the X value represents the amount (or time) of the predictor variable. For the data in Table 1.2, three new predictors can be defined:

X_8 = 1 if from a member of Group One, 2 if from a member of Group 2, 3 if from a member of Group 3, or 4 if from a member of Group 4;

$X_9 = X_8^2$; and $X_{10} = X_8^3$. The reason that three predictors (or implying a third degree polynomial) were defined is that the maximum degree polynomial possible is one less than the number of unique X values; since X takes on the values 1, 2, 3 and 4, the maximum polynomial is a third degree polynomial. For a formulation of the polynomial regression, see Table 3.4.

TABLE 3.4

FORMULATION FOR POLYNOMIAL REGRESSION

Y	X_1	X_2	X_3	X_4	X_8	X_9	X_{10}
9	1	0	0	0	1	1	1
8	1	0	0	0	1	1	1
6	1	0	0	0	1	1	1
3	1	0	0	0	1	1	1
4	1	0	0	0	1	1	1
8	0	1	0	0	2	4	8
7	0	1	0	0	2	4	8
8	0	1	0	0	2	4	8
6	0	1	0	0	2	4	8
6	0	1	0	0	2	4	8
13	0	0	1	0	3	9	27
10	0	0	1	0	3	9	27
12	0	0	1	0	3	9	27
11	0	0	1	0	3	9	27
14	0	0	1	0	3	9	27
15	0	0	0	1	4	16	64
12	0	0	0	1	4	16	64
10	0	0	0	1	4	16	64
17	0	0	0	1	4	16	64
11	0	0	0	1	4	16	64

To accomplish a trend analysis with the data in Table 3.4, a hierarchical solution is necessary. Three successive linear models are solved:

$$Y = b_0 + b_1X + e_3 \quad (3.2)$$

$$Y = b_0 + b_1X + b_2X^2 + e_4 \quad \text{and} \quad (3.3)$$

$$Y = b_0 + b_1X + b_2X^2 + b_3X^3 + e_5. \quad (3.4)$$

Table 3.5 contains the sum of squares, R's and R^2 values for equations 3.2 - 3.4.

TABLE 3.5

TREND ANALYSIS FOR DATA IN TABLE 3.4

Trend	df	SS	MS	F	R	R^2
Linear	1	169.00	169.00	36.541	.80778	.65251
Linear	2	169.00				
Unique to Second	1	0.00	0.00	0.00	.80778	.65251
Linear + Second + Third	3	184.87				
Unique to Third	1	15.87	15.87	3.431	.84486	.71378
Deviation from Full Model	16	74.00	4.625			

The F values in Table 3.5 have been traditionally evaluated by using the F distribution with df = 1, k - 1; this clearly is identical to using the t test ($t = \sqrt{F}$) to evaluate the significance of each component. An alternative process would use the Studentized range statistic with r = 2 for the linear component, r = 3 for the second degree component and r = 4 for the third degree component. Using Table IIb, the linear component is significant ($p < .01$) since $t = \sqrt{36.541} = 6.045 > 2.921$. The second degree component is not significant since $t = \sqrt{0} = 0 < 3.384$. The third degree component is not significant, since $t = \sqrt{3.431} = 1.852 < 3.671$.

One difficulty with the direct use of the X term in polynomial regression can be noted in Table 3.5. The model representing equation 3.4, the third degree polynomial, has an associated sum of squares of 184.87 in Table 3.5. In that a maximum of only a third degree polynomial is possible, the sum of squares <u>should</u> be 185.00. Thus, some computational inaccuracy has entered into the process. Further, if higher degree polynomials are used, the inaccuracies become so gross that they invalidate the analysis. This inaccuracy is somewhat due to the computer program being used, but is primarily due to collinearity. If a correlation matrix of the Y, X, X^2 and X^3 values are investigated, the collinearity becomes obvious.

TABLE 3.6

INTERCORRELATIONS OF Y, X, X^2 and X^3

	Y	X	X^2	X^3
Y		.80778	.79516	.76168
X			.98437	.95137
X^2				.99053
X^3				

One major way of avoiding collinearity is through the use of orthogonal polynomials.

Orthogonal Polynomials

The use of orthogonal polynomials allows a direct solution to the collinearity problem. By having each component orthogonal to all other components, not only is the solution much more accurate than the solution given in the previous section, but the components are additive. Because of the usefulness of orthogonal polynomials, tables of the coefficients are widely available; a table is also reproduced in the Appendix (Table IV).

As there are four groups, the third set of coefficients in Table IV are appropriate. Using these coefficients, three new predictors can be defined:

X_{11} = -3 if a member of Group One, -1 if a member of Group Two, 1 if a member of Group Three, or 3 if a member of Group Four;

X_{12} = 1 if a member of Group One, -1 if a member of Group Two, -1 if a member of Group Three, or 1 if a member of Group Four; and

X_{13} = -1 if a member of Group One, 3 if a member of Group Two, -3 if a member of Group Three, or 1 if a member of Group 4.

While it is not done so here, a table like Table 3.4 could be constructed containing Y, X_{11}, X_{12} and X_{13}. Table 3.7 contains the trend analysis using orthogonal coefficients.

TABLE 3.7

TREND ANALYSIS USING ORTHOGONAL COEFFICIENTS

Trend	df	SS	MS	F	R	R^2
Linear	1	169.00	169.00	36.541	.80778	.65251
Linear + Second	2	169.00			.80778	.65251
Second	1	0.00	0.00	0.00	.00000	.00000
Linear + Second + Third	3	185.00			.84515	.71429
Third	1	16.00	16.00	3.459	.24855	.06178
Deviation from Full Model	16	74.00	4.625			

The results given in Table 3.7 are quite similar to those in Table 3.6. The last F value 3.459 differs somewhat from the corresponding value in Table 3.6 (the SS for Linear + Second + Third and Third also differ); the results in Table 3.7 are the more accurate. Further, any component (Linear or Second or Third) can be found directly (the R and R^2 values respectively for linear, second and third are found directly), since the predictors are orthogonal.

Had there been more than a third degree equation sought, the orthogonal process continues to give exact results, while the direct approach of using $Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_{k-1}X_{k-1} + e$ tends to quickly show computational inaccuracies (the point at which this occurs differs from program to program). On the other hand, researchers prefer to stay away from higher degree polynomials for another reason; using a fourth degree or larger polynomial rarely yields a parsimonious explanation of the data.

Thus, the use of orthogonal coding schemes for either multiple comparisons or trend analysis likely rests with the needs and means of the individual researcher. As a generalized multiple comparison technique, the orthogonalization process will be found lacking by many researchers. Some will find them inconvenient to use because of the difficulty in finding useful orthogonal sets. Others may find that using orthogonal comparisons often leaves untested relevant hypotheses while simultaneously testing orthogonal, but experimentally uninteresting contrasts. Still others would find typically reported error rates to be wanting in comparison to other tests that maintain at least an experimentwise error rate. The use of orthogonal coding can prove to be useful in trend analysis, particularly if higher degree polynomials are sought.

CHAPTER IV

DUNN'S TEST AND SCHEFFE'S TEST

Perhaps one of the most widely disseminated multiple comparison procedures is Scheffé's (1953) S-method for judging all possible contrasts. As Scheffé's test has the property that it can be used on an a posteriori basis after the rejection of the overall F test in the analysis of variance and still maintain an experimentwise error rate, applied researchers find the test to be particularly useful in investigating unplanned contrasts that are suggested by the data.

Dunn (1961) devised a multiple comparison technique that allows the researcher to state the hypotheses to be tested on an a priori basis; if the number of comparisons to be made can be severely limited, Dunn's test can be the most powerful multiple comparison method under these circumstances, and retains a per experiment error rate. Dunn's test would be particularly appropriate in those theory testing situations that require less than all possible simple comparisons of means. A regression formulation of Dunn's test and Scheffé's test was given in Williams (1975).

An Example

In that Dunn's test requires a severe limitation of the number of comparisons of interest, suppose the comparisons of interest when four groups are being included in the analysis are

$\overline{X}_1$ to $\overline{X}_3$,

$\overline{X}_1$ to $\overline{X}_4$, and

$\overline{X}_2$ to $\overline{X}_4$.

Before completing the analysis, comments on other multiple comparison procedures with these specific hypotheses should be made. First, Dunnett's test would not apply, as there is no single group being compared to other groups. Orthogonal comparisons would not apply, as the stated comparisons do not transform to an orthogonal set of comparisons. Both Tukey's and Scheffé's tests would appropriately contain these comparisons (among a much larger set), but with considerably less power. Thus, the kind of experimental situation implied in the posited comparisons fits very well the test proposed by Dunn. In that the three posited comparisons are "simple" in that no more than two means are being compared at any one time, a methodology very similar to that given in Chapter Two can be employed. The process involves employing a pseudo-replication of Dunnett's test until all comparisons are made. As was done previously, the data in Table 1.1 is used. The comparisons of $\overline{X}_1$ to $\overline{X}_4$ and $\overline{X}_1$ to $\overline{X}_3$ can be found from using the linear model

$$Y = b_0 + b_2X_2 + b_3X_3 + b_4X_4 + e_1. \qquad (4.1)$$

The comparison of $\overline{X}_2$ to $\overline{X}_4$ can be found from using the linear model

$$Y = b_0 + b_1X_1 + b_3X_3 + b_4X_4 + e_1. \qquad (4.2)$$

It should be pointed out that equations 4.1 and 4.2 are respectively identical to equations 1.6 and 2.1.

As was true with Tukey's test, the value for b_0 will in general be different for each equation; b_0 will equal the mean of the group that is being compared to the other groups. Also, the regression coefficients are the difference between the means of each group and the group "left out" in a particular analysis. The value for e_1 will remain the same for each of the equations, as each of these equations can be conceptualized as a reparameterization of an analysis of variance model. Using the two linear models, the computed t value for each regression coefficient is found. Each linear model generates three computed t values; however, only three of the six computed t values are of interest.

The computed t values for equation 4.1 are

t = .735 for comparing $\overline{X}_1$ to $\overline{X}_2$ (not a comparison of interest),

t = 4.411 for comparing $\overline{X}_1$ to $\overline{X}_3$, and

t = 5.147 for comparing $\overline{X}_1$ to $\overline{X}_4$.

The computed t values for eauation 4.2 are

t = -.735 for comparing $\overline{X}_2$ to $\overline{X}_1$ (not a comparison of interest),

t = 3.676 for comparing $\overline{X}_2$ to $\overline{X}_3$ (not a comparison of interest),

t = 4.411 for comparing $\overline{X}_2$ to $\overline{X}_4$.

To evaluate these t values, Dunn's (1961) table can be used; Table V a, b, reproduced in the Appendix, are respectively Dunn's tables for the .05 and .01 levels. With 3 comparisons and df_w = 16, the critical value is 3.45 (found by linear interpolation), with α= .01.

TABLE 4.1

COMPUTED t VALUES FOR DUNN'S AND SCHEFFE'S TESTS
IN A REGRESSION FORMULATION

Comparison	t
$\overline{X}_1$ to $\overline{X}_3$	4.411*
$\overline{X}_1$ to $\overline{X}_4$	4.147*
$\overline{X}_2$ to $\overline{X}_4$	4.411*

*Significant at the .01 level

The formulation given for Dunn's test is identical for completing Scheffé's test. The only difference in the interpretation lies in determining the critical value; for Scheffé's test;

$S = \sqrt{(k-1)\,{}_{\alpha}F_{k-1,n-k}}$, which for $\alpha = .01$, $k = 4$, $n = 20$, is $S = 3.98$.

The corresponding critical value for Tukey's test is 3.67. That Dunn's test is more powerful than Scheffé's test for this set of posited simple comparisons is due to the limited number of comparisons. For the application given, then, Dunn's test is the most powerful and Scheffé's test the least powerful, with Tukey's test being intermediate. For the three comparisons given, however, significance at the .01 level for all three comparisons by all three tests was attained.

Complex Comparisons

Both Dunn's test and Scheffé's test allow for more complex comparisons (more typically called contrasts) of the means than is indicated by the three comparisons already shown. The more complex comparisons typically will require a more complex solution than was demonstrated for the simple comparisons.

Suppose the following four comparisons were of interest:

$\frac{1}{2}\overline{X}_1 + \frac{1}{2}\overline{X}_2$ to $\frac{1}{2}\overline{X}_3 + \frac{1}{2}\overline{X}_4$,

$\frac{1}{2}\overline{X}_1 + \frac{1}{2}\overline{X}_2$ to $\overline{X}_4$,

$1/3\overline{X}_1 + 2/3\overline{X}_2$ to $3/7\overline{X}_3 + 4/7\overline{X}_4$ and

$\overline{X}_2$ to $\overline{X}_3$.

There are several approaches to solving for these comparisons. One approach is to define full and restricted models, following the methodology of Ward and Jennings (1973). To consider the first comparison, the full model is identical to any one of the oreviously given equations 4.1 or 4.2 as they all yield the same R^2 vlaue of .71429. Actually, the full model can be thought of as

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e_1. \quad (4.3)$$

In that most computer programs automatically generate a constant term (b_0), equation 4.3 will typically not produce a desired result; if, for examole, b_4X_4 is thought of as being b_0, then one solution would be to use equation 4.1, where b_0 will equal $\overline{X}_4$. The restricted model could then be formed from the restriction implied in the comparison

$$\tfrac{1}{2}b_1 + \tfrac{1}{2}b_2 = \tfrac{1}{2}b_3 + \tfrac{1}{2}b_4.$$

If this restriction is shown in terms of b_1, it is given as

$b_1 = b_3 + b_4 - b_2$. If this restriction is made on the full model (equation 4.3),

$$Y = (b_3 + b_4 - b_2)X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e_2,$$

$$Y = b_2(X_2 - X_1) + b_3(X_3 + X_1) + b_4(X_4 + X_1) + e_2,$$

or

$Y = b_2V_1 + b_3V_2 + b_4V_3 + e_2$, where

V_1 = 1 if a member of Group Two, -1 if a member of Group One, 0 otherwise;

V_2 = 1 if a member of Group Three, 1 if a member of Group One, 0 otherwise;

V_3 = 1 if a member of Group Four, 1 if a member of Group One, 0 otherwise.

Then, a restricted model could be used which uses any two of variables of V_1, V_2 and V_3. Using V_1 and V_2, the restricted model is

$$Y = b_0 + b_5V_1 + b_6V_2 + e_6. \tag{4.4}$$

The resulting R^2 for equation 4.4 is .01931.

To evaluate the significance of this restriction, the equation

$$F = \frac{(R^2_{FM} - R^2_{RM})/1}{(1 - R^2_{FM})/df_w}, \tag{4.5}$$

can be used, where R^2_{FM} refers to the R^2 value from the full model, R^2_{RM} refers to the R^2 value from the restricted model and df_w refers to the degrees of freedom for the within term in the full model.

For the comparison $\frac{1}{2}X_1 + \frac{1}{2}X_2$ to $\frac{1}{2}X_3 + \frac{1}{2}X_4$,

$$F = \frac{(.71429 - .01931)/1}{(1 - .71429)/16} = 38.920.$$

If $\sqrt{F}$ is found, $\sqrt{F} = 6.239$, which can be interpreted as a t value for Dunn's test (or as the value of S in Scheffé's test) for this comparison.

A simpler approach can be used to gain the same result. If a new predictor V_4 is defined as $V_4 = \frac{1}{2}$ for a member of Group One, $\frac{1}{2}$ for a member of Group Two, $-\frac{1}{2}$ for a member of Group Three or $-\frac{1}{2}$ for a member of Group Four, so that V_4 is a direct utilization of the comparison of interest, then the R^2 value found from using V_4 as a predictor is .69498, which is equal to $R^2_{FM} - R^2_{RM}$: $.71429 - .01931 = .69498$. The value of R^2 with the use of V_4 can be directly utilized as equal to $R^2_{FM} - R^2_{RM}$ in equation 4.5. Similar solutions can be given for the comparisons

$\frac{1}{2}\overline{X}_1 + \frac{1}{2}\overline{X}_2$ to $\overline{X}_4$ ($F = 30.451$, $t = 5.518$),

$1/3\overline{X}_1 + 2/3\overline{X}_2$ to $3/7\overline{X}_3 + 4/7\overline{X}_4$ ($F = 35.368$, $t = 5.947$) and

$\overline{X}_2$ to $\overline{X}_3$ ($F = 13.514$, $t = 3.676$).

The last comparison is of particular interest in that it is of the type given earlier for simple comparisons and is in fact identically equal to the earlier obtained result from the value given from the use of equation 4.2, although it was not then a comparison of interest.

To evaluate these comparisons for significance, the tabled (using interpolation) value for Dunn's test at the .01 level for four comparisons is 3.59. Each of the four comparisons is significant. The critical value for Scheffé's test remains as 3.98. While Tukey's test also yields significance on the four posited tests, it is of interest to note that on each of the three complex comparisons, Scheffé's test has a shorter interval than does Tukey's test.

As was noticed, including four comparisons rather than three reduced somewhat the advantages of Dunn's test over Scheffé's test. For $\alpha = .01$, $k = 4$, $df_w = 16$, Dunn's test will be more powerful than Scheffé's test for any number of comparisons up to 9 comparisons. Beyond 9 comparisons, Scheffé's test is more powerful. As both Dunn's and Scheffé's test employ the same standard error for a contrast, they are easily comparable. On the other hand, comparisons to Tukey's test are more difficult.

If there is interest in finding complex comparisons using Tukey's test, and the following conventions hold: $\Sigma a_i = 0$, $\Sigma|a_i| = 2$, so that the sum of the positive contrast coefficients is 1 and the sum of the negative contrast coefficients is -1, then a transformation of the computed t value can be made so that Tables IIa, b can be used directly. The transformation is given by

$$\text{Tukey's } t = \sqrt{\left(\frac{\Sigma a_i^2}{2}\right)}\, t_c \qquad (4.6)$$

where t_c is the computed t value. For the complex comparison $\frac{1}{2}\bar{X}_1 + \frac{1}{2}\bar{X}_2$ to $\frac{1}{2}\bar{X}_3 + \frac{1}{2}\bar{X}_4$, the contrast can be written

$d_1 = ½\overline{X}_1 + ½\overline{X}_2 - ½\overline{X}_3 - ½\overline{X}_4$. Then $\Sigma a_i^2 = ½^2 + ½^2 + (-½)^2 + (-½)^2 = 1$.

$$\text{Tukey's } t = \sqrt{½}\, t_c$$
$$= (.7071)(6.239) = 4.411.$$

This same Tukey t value was found on two occasions in Chapter Two. The comparisons were $\overline{X}_1$ to $\overline{X}_3$ and $\overline{X}_2$ to $\overline{X}_4$. In both those cases, the differences in the means is 6.0. In the present contrast, the difference in the means is also 6.0, since ½(6.0) + ½(7.0)-½(12.0) - ½(13.0) = -6.0.

For the other three contrasts, Tukey's t values are:

$$t = \{\sqrt{½\left[(½)^2 + (½)^2 + 1^2\right]}\}\ (5.518) = 4.779$$

for the contrast $d_2 = ½\overline{X}_1 + ½\overline{X}_2 - \overline{X}_4$;

$$t = \{\sqrt{½\left[(1/3)^2 + (2/3)^2 + (3/7)^2 + (4/7)^2\right]}\}(5.947) = 4.341$$

for the contrast $d_3 = 1/3\overline{X}_1 + 2/3\overline{X}_2 - 3/7\overline{X}_3 - 4/7\overline{X}_4$; and

$$t = \{\sqrt{½\ (1^2 + 1^2)}\ \}\ (3.676) = 3.676$$

for the contrast $d_4 = \overline{X}_2 - \overline{X}_3$.

Tukey's test maximizes its power with simple comparisons of the form $\overline{X}_i - \overline{X}_j$; for complex comparisons Scheffé's test tends to be more powerful.

If all seven of the previously given comparisons had been posited a priori, then Dunn's test has a critical value of 3.86, with Scheffé's test critical value remaining at 3.98. For such a situation, Tukey's test would be more powerful for the four simple comparisons, but both Dunn's and Scheffé's tests are more powerful on the complex comparisons, with Dunn's test being slightly the more powerful of the two.

Multiple Linear Regression Viewpoints
Vol. 7, No. 1, 1976
Monograph Series #2

CHAPTER V

FINDING THE MAXIMIZED SCHEFFE CONTRAST THROUGH MULTIPLE REGRESSION

Scheffé has indicated that whenever an overall F test allows the rejection of the overall hypothesis, then at least one linear contrast will be significant; this, however, has to many researchers seemed to be an empty promise; how do you find that significant contrast? The present chapter is directed to the solution to the just posed question.

As was true of the previous chapters, the data in Table 1.1 is used. While the data does include some simple comparisons that would yield significance, the present focus is on finding the most powerful contrast. Once that contrast is found, comparisons may suggest themselves that will be "close to" the maximized contrast.

If standardized regression coefficients are employed in the solution to the data in Table 1.1, an equation like

$$z_y = \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_{k-1} z_{k-1} + e_7 \tag{5.1}$$

results. In equation 5.1 the data has first been standardized so that each variable has a mean of 0 and a standard deviation of 1. Using the beta coefficients given in Table 1.3,

$$z_y = .12033z_2 + .72197z_3 + .84230z_4 + e_8. \tag{5.2}$$

If the predicted values for z_y are found, the resulting values will constitute the coefficients for a maximized Scheffé contrast. To find z_2, z_3 and z_4, the mean for X_2, X_3 and X_4 are available from the printout given in Table 1.3 and all three variables have a mean of .25. In this case, the mean is simply the proportion of scores in each group. Also, the standard deviation for each variable is given as .44426. This number might seem somewhat synthetic as a standard deviation, since five of the scores are 1 and the remaining 15 scores are 0 for variables X_2, X_3 and X_4. However, the standard deviation is useful in a computational sense. Then

$$z_2 = \frac{X_2 - .25}{.44426}$$

Similarly,

$$z_3 = \frac{X_3 - .25}{.44426}$$

and

$$z_4 = \frac{X_4 - .25}{.44426}$$

Equation 5.2 can be rewritten as

$$z_y = .12033^* \left[(X_2 - .25)/.44426\right] + .72197^* \left[(X_3 - .25)/.44426\right] + .84230^* \left[(X_4 - .25)/.44426\right]. \tag{5.3}$$

If equation 5.3 is solved for a member of Group One,

$$z_{y_1} = .12033^* \left[(0-.25)/.44426\right] + .72197^* \left[(0 - .25)/.44426\right] + .84230^* \left[(0 - .25)/.44426\right] = -.94798.$$

In a similar manner, the predicted values of z_y can be found for members of the other three groups:

V_5= -.94798 for members of Group 1;

-.67713 for members of Group 2;

.67713 for members of Group 3;

.94798 for members of Group 4.

Then a contrast can be defined as

$$\Psi_1 = -.94798X_1 - .67713X_2 + .67713X_3 + .94798X_4 . \tag{5.4}$$

The linear model for this contrast is

$$Y = b_0 + b_1V_5 + e_1. \tag{5.5}$$

The use of the contrast given by equation 5.4 in the linear model in equation 5.5 also yields

$$R = .84515$$

and

$$R^2 = .71429.$$

In other words, the <u>entire</u> difference among the four groups is expressed by the contrast in equation 5.4. Thus, if the F test is signfficant, this contrast will also be significant. While the example given here had equal n in each group, equal n are not necessary to find the maximum Scheffé contrast.

When using unequal n, it is necessary to use the mean and standard deviation for a given group, where the mean is the proportion of subjects in that group, and the standard deviation is a standard deviation for this proportion. These values typically accompany most general purpose multiple regression programs.

The choice of X_2 - X_4 was arbitrary in the sense that any three of the four group membership variables will yield the same coefficients as is indicated for equation 5.4. The contrast found by this process is unique up to a multiplication by a constant.

In that <u>any</u> contrast that is proportional to Ψ_1 will yield an R^2 value identical to the R^2 found in the analysis of variance, the researcher is free to fix at any value (other than zero) a given weighting and weight the remaining groups. For example, some researchers may prefer to take the largest (in absolute value) weighting and divide all other weights by this weighting (in absolute value). For example, if all weights in Ψ_1 are divided by .94798, then the derived contrast is

$$\Psi_2 = -1.00X_1 - .71X_2 + .71X_3 + 1.00X_4 \text{ (rounded to two decimal points);}$$

Ψ_2 also yields

$R = .84515$ and

$R^2 = .71429$

but has the advantage of being somewhat more interpretable than Ψ_1.

Multiple Linear Regression Viewpoints
Vol. 7, No. 1, 1976
Monograph Series #2

CHAPTER VI

ON CHOOSING A MULTIPLE COMPARISON METHOD

An attempt was made to present the various multiple comparison methods in their best light in Chapters I-V; few judgments regarding their usefulness are made therein. In the present chapter, an attempt is made to indicate the domain of usefulness of the various tests; some value judgments are included.

In deciding which test to employ, several questions should be answered:

1. Is the researcher willing to state the contrasts of interest <u>before</u> the data is collected?
2. Are simple comparisons of means of highest importance, or do some complex comparisons (contrasts) have equal (or higher) research interest?
3. What kind of error rate is viewed as being appropriate?

If the answer to question 1 (beforehand contrasts) is <u>yes</u>, then, of course, the researcher should state them. If the stated contrasts are <u>exactly</u> comparing one group to all other groups, then Dunnett's test is appropriate. If the number of comparisons is small, but not identical to those tested by Dunnett's test, then either Dunn's test or orthogonal comparisons might be used; orthogonal comparisons are appropriate <u>only</u> if the comparisons stated <u>a priori</u> are orthogonal (a seemingly unlikely event); then, of course, the orthogonal comparison procedure can be used <u>provided</u> the error rate is adjusted to being experimentwise (an admitted value judgment). Otherwise, Dunn's test is more appropriate.

If the answer to question 1 is <u>no</u>, then question 2 should be answered. If the number of comparisons (contrasts) greatly exceeds $2\binom{k}{2}$ where k is the number of groups, then Scheffè's test should probably be used, particularly if the complex comparisons are of major research interest. Even if the number of comparisons does not exceed $2\binom{k}{2}$, if the majority of contrasts of interest are complex, Scheffé's test would still have the highest probability of rejecting a null hypothesis on an a posteriori basis and still retain an experimentwise error rate.

If the comparisons of interest form an exact set of orthogonal contrasts, then the orthogonal comparisons might be used on an a posteriori basis. (Some might dispute this usage, preferring that the orthogonal contrasts be used only on an a priori basis).

If only the simple comparisons of the means are of interest, and if all simple comparisons are being pursued on an a posteriori basis, then the methods considered in Chapter IV are appropriate: Tukey's (a) H.S.D. test, Tukey's (b) test, Newman-Keul's test and Duncan's test. Choosing among these four tests depends upon how one views probability. Clearly, Tukey's (a) H.S.D. test is appropriate, but is not as powerful as the other methods. So that the reader is forewarned, the present writer is going to take a stand against the other three tests. Duncan's approach to fixing the level of an experiment is particularly perplexing. If Duncan's test reports an alpha of .05, one might ask, 5% of what? The alpha level does not represent a probability value (except in the trivial case when $k = 2$); it would appear that many users of the Duncan method do so without understanding the limitations of the Duncan "protection levels". The Newman-Keuls procedure (and by inference, the Tukey (b) test) implictly requires at least a weak ordering of the means in the population to correspond to the sample findings. If the sample findings are $\overline{X}_1 < \overline{X}_2 < \overline{X}_3 < \overline{X}_4$, then an implicit assumption is that $\mu_1 \leq \mu_2 \leq \mu_3 \leq \mu_4$.

The additional cost associated with the Tukey (a) H.S.D. test seems worth the dropping of this assumption.

Thus, a recommendation is made that users of regression become familiar with at least the Tukey (a) H.S.D, Dunnett's, Dunn's and Scheffé's tests. Additionally, the orthogonal methodology may occasionally be useful for multiple comparisons but will likely be useful in situations dealing with trends. A point that bears emphasis is that when multiple tests are being made on the same full model (such as occurs in the multiple comparison situation), the probability should be adjusted in keeping with the number of tests.

As an example, suppose three a priori restricted models are imposed on the same full model. To use the F table to evaluate each of these outcomes separately necessarily takes advantage of chance. One simple (but conservative) way to extricate the taking advantage of chance is to multiply each probability found by the number of tests performed. If the three probabilities found were $p_1 = .0027$, $p_2 = .0100$, $p_3 = .04$, then the reported probabilities would be $3p_1 = 3(.0027) = .0081$, $3p_2 = .03$; and $3p_3 = .12$. While such a change might not be popular with researchers, the goal of science is to achieve stability in research findings, not to report chance events (and perhaps massage the ego of a "lucky" researcher). The adjustment of the probabilities through either the multiple comparison methods (if appropriate) or by multiplying the obtained probabilities by the number of tests performed (thereby achieving a per experiment error rate) is clearly preferable to reporting chance findings as significant.

While the preceding suggestions implicitly assume that those using multiple linear regression have been neglectful of the effect on probability when making multiple use of the same full model without taking into account the multiple usage, other researchers have shown concern as well. Connett (1971) and Newman and Fry (1972) have presented tables for Kimball's (1951) method for keeping α levels constant when making a number of comparisons. In a related vein, Newman <u>et al.</u> (1976) have considered the "Type VI error" in regression. A Type VI error is an inconsistency between the researcher's question of interest and the statistical procedures employed to analyze the data. While it might be presumptuous to label the overusage of a single full model without adjusting the probability level as a "Type VII error," it is tempting.

TABLES

Tables Ia, Ib, Ic, and Id are reproduced from C. W. Dunnett, A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association, 1955, 50, 1096-1122, and C. W. Dunnett, New tables for multiple comparisons with a control. Biometrics, 1964, 20, 482-491, with permission of the author.

Tables IIa and IIb were calculated by the present writer by transforming the values in Harter (1960) by division by $\sqrt{2}$.

Tables IIIa and IIIb were calculated by the present writer by transforming values in Duncan (1955) by division by $\sqrt{2}$.

Table IV was taken from Anderson and Houseman (1942), by permission of first author.

Tables Va and Vb are reproduced from O. J. Dunn, Multiple comparisons among means. Journal of the American Statistical Association, 1961, 56: 52-64 by permission of the author.

In Tables Ia, Ib, Ic, Id, IIa, IIb, IIIa, IIIb, Va, and Vb, v refers to the degrees of freedom within (df_w).

In Tables IIa, IIb, IIIa, IIIb, r refers to the number of means in the range.

In Tables Ia, Ib, Ic, Id, k refers to the number of groups compared to the control (excluding the control).

In Tables Va and Vb, m refers to the total number of *a priori* contrasts.

TABLE Ia

PERCENTAGE POINTS OF DUNNETT'S TEST
(.05 LEVEL)

One-tailed

v \ k	1	2	3	4	5	6	7	8	9
5	2.02	2.44	2.68	2.85	2.98	3.08	3.16	3.24	3.30
6	1.94	2.34	2.56	2.71	2.83	2.92	3.00	3.07	3.12
7	1.89	2.27	2.48	2.62	2.73	2.82	2.89	2.95	3.01
8	1.86	2.22	2.42	2.55	2.66	2.74	2.81	2.87	2.92
9	1.83	2.18	2.37	2.50	2.60	2.68	2.75	2.81	2.86
10	1.81	2.15	2.34	2.47	2.56	2.64	2.70	2.76	2.81
11	1.80	2.13	2.31	2.44	2.53	2.60	2.67	2.72	2.77
12	1.78	2.11	2.29	2.41	2.50	2.58	2.64	2.69	2.74
13	1.77	2.09	2.27	2.39	2.48	2.55	2.61	2.66	2.71
14	1.76	2.08	2.25	2.37	2.46	2.53	2.59	2.64	2.69
15	1.75	2.07	2.24	2.36	2.44	2.51	2.57	2.62	2.67
16	1.75	2.06	2.23	2.34	2.43	2.50	2.56	2.61	2.65
17	1.74	2.05	2.22	2.33	2.42	2.49	2.54	2.59	2.64
18	1.73	2.04	2.21	2.32	2.41	2.48	2.53	2.58	2.62
19	1.73	2.03	2.20	2.31	2.40	2.47	2.52	2.57	2.61
20	1.72	2.03	2.19	2.30	2.39	2.46	2.51	2.56	2.60
30	1.70	1.99	2.15	2.25	2.33	2.40	2.45	2.50	2.54
40	1.68	1.97	2.13	2.23	2.31	2.37	2.42	2.47	2.51
60	1.67	1.95	2.10	2.21	2.28	2.35	2.39	2.44	2.48
120	1.66	1.93	2.08	2.18	2.26	2.32	2.37	2.41	2.45
∞	1.64	1.92	2.06	2.16	2.23	2.29	2.34	2.38	2.42

TABLE Ib

PERCENTAGE POINTS OF DUNNETT'S TEST
(.01 LEVEL)

One-tailed

v \ k	1	2	3	4	5	6	7	8	9
5	3.37	3.90	4.21	4.43	4.60	4.73	4.85	4.94	5.03
6	3.14	3.61	3.88	4.07	4.21	4.33	4.43	4.51	4.59
7	3.00	3.42	3.66	3.83	3.96	4.07	4.15	4.23	4.30
8	2.90	3.29	3.51	3.67	3.79	3.88	3.96	4.03	4.09
9	2.82	3.19	3.40	3.55	3.66	3.75	3.83	3.89	3.94
10	2.76	3.11	3.31	3.45	3.56	3.64	3.71	3.78	3.83
11	2.72	3.06	3.25	3.38	3.48	3.56	3.63	3.69	3.74
12	2.68	3.01	3.19	3.32	3.42	3.50	3.56	3.62	3.67
13	2.65	2.97	3.15	3.27	3.37	3.44	3.51	3.56	3.61
14	2.62	2.94	3.11	3.23	3.32	3.40	3.46	3.51	3.56
15	2.60	2.91	3.08	3.20	3.29	3.36	3.42	3.47	3.52
16	2.58	2.88	3.05	3.17	3.26	3.33	3.39	3.44	3.48
17	2.57	2.86	3.03	3.14	3.23	3.30	3.36	3.41	3.45
18	2.55	2.84	3.01	3.12	3.21	3.27	3.33	3.38	3.42
19	2.54	2.83	2.99	3.10	3.18	3.25	3.31	3.36	3.40
20	2.53	2.81	2.97	3.08	2.17	3.23	3.29	3.34	3.38
30	2.46	2.72	2.87	2.97	3.05	3.11	3.16	3.21	3.24
40	2.42	2.68	2.82	2.92	2.99	3.05	3.10	3.14	3.18
60	2.39	2.64	2.78	2.87	2.94	3.00	3.04	3.08	3.12
120	2.36	2.60	2.73	2.82	2.89	2.94	2.99	3.03	3.06
∞	2.33	2.56	2.68	2.77	2.84	2.89	2.93	2.97	3.00

TABLE Ic

PERCENTAGE POINTS OF DUNNETT'S TEST
(.05 LEVEL)

Two-tailed

v \ k	1	2	3	4	5	6	7	8	9
5	2.57	3.03	3.29	3.48	3.69	3.73	3.82	3.90	3.97
6	2.45	2.86	3.10	3.26	3.39	3.49	3.57	3.64	3.71
7	2.36	2.75	2.97	3.12	3.24	3.33	3.41	3.47	3.53
8	2.31	2.67	2.88	3.02	3.13	3.22	3.29	3.5	3.41
9	2.26	2.61	2.81	2.95	3.05	3.14	3.20	3.26	3.32
10	2.23	2.57	2.76	2.89	2.99	3.07	3.14	3.19	3.24
11	2.20	2.53	2.72	2.84	2.94	3.02	3.08	3.14	3.19
12	2.18	2.50	2.68	2.81	2.90	2.98	3.04	3.09	3.14
13	2.16	2.48	2.65	2.78	2.87	2.94	3.00	3.06	3.10
14	2.14	2.46	2.63	2.75	2.84	2.91	2.97	3.02	3.07
15	2.13	2.44	2.61	2.73	2.82	2.89	2.95	3.00	3.04
16	2.12	2.42	2.59	2.71	2.80	2.87	2.92	2.97	3.02
17	2.11	2.41	2.58	2.69	2.78	2.85	2.90	2.95	3.00
18	2.10	2.40	2.56	2.68	2.76	2.83	2.89	2.94	2.98
19	2.09	2.39	2.55	2.66	2.75	2.81	2.87	2.92	2.96
20	2.09	2.38	2.54	2.56	2.73	2.80	2.86	2.90	2.95
30	2.04	2.32	2.47	2.58	2.66	2.72	2.77	2.82	2.86
40	2.02	2.29	2.44	2.54	2.62	2.68	2.73	2.77	2.81
60	2.00	2.27	2.41	2.51	2.58	2.64	2.69	2.73	2.77
120	1.98	2.24	2.38	2.47	2.55	2.60	2.65	2.69	2.73
∞	1.96	2.21	2.35	2.44	2.51	2.56	2.61	2.65	2.69

TABLE Id

PERCENTAGE POINTS OF DUNNETT'S TEST
(.01 LEVEL)

Two-tailed

v \ k	1	2	3	4	5	6	7	8	9
5	4.03	4.63	4.98	5.22	5.41	5.56	5.69	5.80	5.89
6	3.71	4.21	4.51	4.71	4.87	5.00	5.10	5.20	5.28
7	3.50	3.95	4.21	4.39	4.53	4.64	4.74	4.82	4.89
8	3.36	3.77	4.00	4.17	4.29	4.40	4.48	4.56	4.62
9	3.25	3.63	3.85	4.01	4.12	4.22	4.30	4.37	4.43
10	3.17	3.53	3.74	3.88	3.99	4.08	4.16	4.22	4.28
11	3.11	3.45	3.65	3.79	3.89	3.98	4.05	4.11	4.16
12	3.05	3.39	3.58	3.71	3.81	3.89	3.95	4.02	4.07
13	3.01	3.33	3.52	3.65	3.74	3.82	3.89	3.94	3.99
14	2.98	3.29	3.47	3.59	3.69	3.76	3.83	3.88	3.93
15	2.95	3.25	3.43	3.55	3.64	3.71	3.78	3.83	3.88
16	2.92	3.22	3.39	3.51	3.60	3.67	3.73	3.78	3.83
17	2.90	3.19	3.36	3.47	3.56	3.63	3.69	3.74	3.79
18	2.88	3.17	3.33	3.44	3.53	3.60	3.66	3.71	3.75
19	2.86	3.15	3.31	3.42	3.50	3.56	3.63	3.68	3.72
20	2.85	3.13	3.29	3.40	3.48	3.55	3.60	3.65	3.69
30	2.75	3.01	3.15	3.25	3.33	3.39	3.44	3.49	3.52
40	2.70	2.95	3.09	3.19	3.26	3.32	3.37	3.41	3.44
60	2.66	2.90	3.03	3.12	3.19	3.25	3.29	3.33	3.37
120	2.62	2.85	2.97	3.06	3.12	3.18	3.22	3.26	3.29
∞	2.58	2.79	2.92	3.00	3.06	3.11	3.15	3.19	3.22

TABLE IIa

PERCENTAGE POINTS OF THE STUDENTIZED RANGE
REPORTED AS t VALUES (.05 LEVEL)

v \ r	2	3	4	5	6	7	8	9	10	20
5	2.470	3.254	3.690	4.011	4.266	4.476	4.654	4.810	4.946	5.804
6	2.447	3.068	3.462	3.751	3.980	4.168	4.329	4.468	4.591	5.365
7	2.365	2.945	3.310	3.578	3.789	3.964	4.112	4.241	4.354	5.070
8	2.306	2.857	3.202	3.455	3.654	3.818	3.958	4.078	4.185	4.858
9	2.262	2.792	3.122	3.363	3.553	3.709	3.841	3.956	4.058	4.698
10	2.228	2.749	3.060	3.291	3.473	3.623	3.751	3.862	3.959	4.573
11	2.201	2.701	3.009	3.234	3.410	3.555	3.678	3.785	3.880	4.473
12	2.179	2.668	2.969	3.188	3.359	3.500	3.620	3.723	3.815	4.390
13	2.160	2.641	2.935	3.149	3.316	3.454	3.550	3.671	3.760	4.322
14	2.145	2.618	2.907	3.116	3.280	3.415	3.528	3.628	3.715	4.263
15	2.131	2.598	2.882	3.088	3.249	3.381	3.493	3.590	3.676	4.213
16	2.120	2.580	2.861	3.064	3.222	3.352	3.463	3.557	3.642	4.170
17	2.110	2.565	2.843	3.043	3.199	3.327	3.435	3.529	3.612	4.131
18	2.101	2.553	2.826	3.024	3.178	3.304	3.411	3.504	3.586	4.097
19	2.093	2.541	2.812	3.007	3.160	3.285	3.390	3.482	3.562	4.067
20	2.086	2.530	2.799	2.992	3.143	3.267	3.371	3.462	3.541	4.040
30	2.042	2.465	2.719	2.913	3.042	3.157	3.254	3.338	3.411	3.871
40	2.021	2.434	2.681	2.856	2.992	3.103	3.197	3.277	3.348	3.789
60	2.071	2.403	2.642	2.812	2.944	3.050	3.140	3.217	3.285	3.706
120	1.980	2.373	2.606	2.770	2.896	2.999	3.085	3.159	3.224	3.625
∞	1.960	2.343	2.569	2.727	2.850	2.949	3.031	3.102	3.164	3.544

TABLE IIb

PERCENTAGE POINTS OF THE STUDENTIZED RANGE
REPORTED AS t VALUES (.01 LEVEL)

v \ r	2	3	4	5	6	7	8	9	10	20
5	4.032	4.933	5.518	5.955	6.302	6.591	6.837	7.051	7.241	8.456
6	3.707	4.476	4.973	5.343	5.638	5.882	6.909	6.271	6.433	7.453
7	3.499	4.185	4.627	4.953	5.213	5.430	5.614	5.774	5.917	6.821
8	3.356	3.985	4.387	4.685	4.921	5.117	5.285	5.431	5.560	6.383
9	3.259	3.838	4.212	4.489	4.708	4.890	5.044	5.180	5.300	6.062
10	3.169	3.726	4.079	4.339	4.545	4.716	4.861	4.953	5.100	5.817
11	3.106	3.639	3.975	4.221	4.417	4.579	4.718	4.838	4.944	5.623
12	3.055	3.568	3.891	4.127	4.314	4.470	4.601	4.716	4.818	5.467
13	3.012	3.510	3.821	4.050	4.229	4.378	4.474	4.616	4.714	5.337
14	2.977	3.461	3.763	3.982	4.158	4.303	4.425	4.532	4.627	5.229
15	2.947	3.420	3.714	3.929	4.098	4.238	4.357	4.461	4.553	5.136
16	2.921	3.384	3.671	3.881	4.046	4.183	4.299	4.400	4.489	5.057
17	2.898	3.353	3.635	3.840	4.002	4.134	4.248	4.347	4.434	4.987
18	2.879	3.326	3.602	3.804	3.962	4.093	4.203	4.300	4.385	4.927
19	2.861	3.302	3.574	3.772	3.927	4.055	4.164	4.258	4.342	4.873
20	2.845	3.280	3.548	3.743	3.896	4.022	4.129	4.221	4.304	4.825
30	2.750	3.150	3.393	3.569	3.707	3.819	3.915	3.997	4.070	4.530
40	2.705	3.088	3.321	3.487	3.616	3.723	3.813	3.891	3.959	4.390
60	2.660	3.028	3.249	3.407	3.529	3.630	3.714	3.787	3.852	4.253
120	2.618	2.970	3.180	3.330	3.445	3.539	3.619	3.687	3.747	4.120
∞	2.576	2.913	3.113	3.255	3.364	3.452	3.526	3.591	3.647	3.992

TABLE IIIa

PERCENTAGE POINTS OF DUNCAN'S MULTIPLE RANGE TEST
REPORTED AS t VALUES (.05 LEVEL)

v \ r	2	3	4	5	6	7	8	9	10	20
5	2.570	2.651	2.685	2.697	2.697	2.697	2.697	2.697	2.697	2.697
6	2.447	2.536	2.580	2.602	2.612	2.614	2.614	2.614	2.614	2.614
7	2.365	2.459	2.509	2.537	2.553	2.561	2.564	2.564	2.564	2.564
8	2.306	2.403	2.457	2.490	2.510	2.522	2.528	2.531	2.531	2.531
9	2.262	2.361	2.418	2.454	2.476	2.491	2.500	2.506	2.508	2.508
10										
10	2.228	2.329	2.387	2.425	2.450	2.469	2.478	2.486	2.490	2.493
11	2.201	2.302	2.363	2.402	2.429	2.448	2.461	2.470	2.476	2.482
12	2.179	2.280	2.343	2.383	2.411	2.432	2.446	2.456	2.462	2.474
13	2.160	2.263	2.326	2.367	2.396	2.418	2.434	2.445	2.453	2.468
14	2.145	2.247	2.311	2.354	2.384	2.406	2.423	2.435	2.444	2.464
15	2.131	2.234	2.298	2.342	2.373	2.396	2.413	2.427	2.437	2.461
16	2.120	2.223	2.287	2.332	2.364	2.387	2.406	2.420	2.430	2.459
17	2.110	2.213	2.278	2.323	2.355	2.380	2.399	2.413	2.425	2.458
18	2.101	2.205	2.270	2.315	2.345	2.373	2.392	2.408	2.419	2.456
19	2.093	2.197	2.262	2.308	2.341	2.367	2.386	3.402	2.415	2.456
20	2.086	2.190	2.256	2.302	2.336	2.361	2.382	2.398	2.411	2.456
30	2.042	2.146	2.214	2.262	2.298	2.326	2.349	2.368	2.384	2.454
40	2.021	2.126	2.193	2.242	2.280	2.309	2.333	2.353	2.370	2.453
60	2.000	2.104	2.173	2.222	2.261	2.292	2.317	2.338	2.357	2.452
120	1.980	2.084	2.153	2.203	2.243	2.275	2.301	2.324	2.343	2.451
∞	1.960	2.063	2.133	2.184	2.225	2.258	2.285	2.309	2.329	2.451

TABLE IIIb

PERCENTAGE POINTS OF DUNCAN'S MULTIPLE RANGE TEST
REPORTED AS t VALUES (.01 LEVEL)

v \ r	2	3	4	5	6	7	8	9	10	20
5	4.032	4.167	4.235	4.271	4.289	4.295	4.295	4.295	4.295	4.295
6	3.707	3.846	3.924	3.970	3.999	4.016	4.026	4.031	4.033	4.033
7	3.499	3.638	3.719	3.772	3.806	3.830	3.846	3.857	3.864	3.870
8	3.356	3.492	3.576	3.631	3.669	3.696	3.717	3.731	3.741	3.760
9	3.250	3.385	3.469	3.526	3.566	3.596	3.619	3.636	3.649	3.681
10	3.169	3.303	3.387	3.444	3.487	3.518	3.543	3.562	3.577	3.623
11	3.106	3.238	3.321	3.380	3.423	3.456	3.482	3.502	3.518	3.577
12	3.055	3.185	3.268	3.328	3.371	3.405	3.431	3.453	3.470	3.540
13	3.012	3.141	3.224	3.284	3.328	3.362	3.389	3.411	3.430	3.507
14	1.977	3.105	3.188	3.246	3.291	3.326	3.354	3.366	3.396	3.480
15	2.947	3.074	3.156	3.215	3.260	2.295	3.323	3.347	3.366	3.456
16	2.921	3.047	3.129	3.188	3.233	3.268	3.297	3.321	3.340	3.435
17	2.898	3.023	3.105	3.164	3.210	3.251	3.274	3.298	3.318	3.417
18	2.879	3.002	3.084	3.143	3.188	3.224	3.253	3.277	3.298	3.400
19	2.861	2.984	3.075	3.125	3.170	3.206	3.235	3.260	3.280	3.386
20	2.845	2.968	3.049	3.108	3.153	3.189	3.219	3.243	3.265	3.372
30	2.750	2.868	2.947	3.005	3.050	3.066	3.118	3.943	3.166	3.288
40	2.705	3.820	2.898	2.956	3.001	3.038	3.068	3.094	3.117	3.246
60	2.660	2.773	2.850	2.907	2.951	2.988	3.019	3.046	3.069	3.203
120	2.618	2.728	2.804	2.860	2.904	2.940	2.971	2.997	3.021	3.160
∞	2.576	2.684	7.758	2.813	2.857	2.893	2.924	2.950	2.973	3.117

TABLE IV

COEFFICIENTS OF ORTHOGONAL POLYNOMIALS

k	Polynomial	X = 1	2	3	4	5	6	7	8	9	10
3	Linear	-1	0	1							
	Quadratic	1	-2	1							
4	Linear	-3	-1	1	3						
	Quadratic	1	-1	-1	1						
	Cubic	-1	3	-3	1						
	Linear	-2	-1	0	1	2					
5	Quadratic	2	-1	-2	-1	2					
	Cubic	-1	2	0	-2	1					
	Quartic	1	-4	6	-4	1					
	Linear	-5	-3	-1	1	3	5				
	Quadratic	5	-1	-4	-4	-1	5				
	Cubic	-5	7	4	-4	-7	5				
	Quartic	1	-3	2	2	-3	1				
	Linear	-3	-2	-1	0	1	2	3			
7	Quadratic	5	0	-3	-4	-3	0	5			
	Cubic	-1	1	1	0	-1	-1	1			
	Quartic	3	-7	1	6	1	-7	3			
	Linear	-7	-5	-3	-1	1	3	5	7		
	Quadratic	7	1	-3	-5	-5	-3	1	7		
8	Cubic	-7	5	7	3	-3	-7	-5	7		
	Quartic	7	-13	-3	9	-9	-3	-13	7		
	Quintic	-7	23	-17	-15	15	17	-23	7		
	Linear	-4	-3	-2	-1	0	1	2	3	4	
	Quadratic	23	7	-8	-17	-20	-17	-8	7	28	
9	Cubic	-14	7	13	9	0	-9	-13	-7	14	
	Quartic	14	-21	-11	9	18	9	-11	-21	14	
	Quintic	-4	11	-4	-9	0	9	4	-11	4	
	Linear	-9	-7	-5	-3	-1	1	3	5	7	9
	Quadratic	6	2	-1	-3	-4	-4	-3	-1	2	6
10	Cubic	-42	14	35	31	12	-12	-31	-35	-14	42
	Quartic	18	-22	-17	3	18	18	3	-17	-22	18
	Quintic	-6	14	-1	-11	-6	6	11	1	-14	6

TABLE Va

VALUES FOR DUNN'S TEST
(.05 LEVEL)

v \ m	2	3	4	5	6	7	8	9	10	20
5	3.17	3.54	3.81	4.04	4.22	4.38	4.53	4.66	4.78	5.60
7	2.84	3.13	3.34	3.50	3.64	3.76	3.86	3.95	4.03	4.59
10	2.64	2.87	3.04	3.17	3.28	3.37	3.45	3.52	3.58	4.01
12	2.56	2.78	2.94	3.06	3.15	3.24	3.31	3.37	3.43	3.80
15	2.49	2.69	2.84	2.95	3.04	3.11	3.18	3.24	3.29	3.62
20	2.42	2.61	2.75	2.85	2.93	3.00	3.06	3.11	3.16	3.46
24	2.39	2.58	2.70	2.80	2.88	2.94	3.00	3.05	3.09	3.38
30	2.36	2.54	2.66	2.75	2.83	2.89	2.94	2.99	3.03	3.30
40	2.33	2.50	2.62	2.71	2.78	2.84	2.89	2.93	2.97	3.23
60	2.30	2.47	2.58	2.66	2.73	2.79	2.84	2.88	2.92	3.16
120	2.27	2.43	2.54	2.62	2.68	2.74	2.79	2.83	2.86	3.09
∞	2.24	2.39	2.50	2.58	2.64	2.69	2.74	2.77	2.81	3.02

TABLE Vb

VALUES FOR DUNN'S TEST
(.01 LEVEL)

v \ m	2	3	4	5	6	7	8	9	10	20
5	4.78	5.25	5.60	5.89	6.15	6.36	6.56	6.70	6.86	8.00
7	4.03	4.36	4.59	4.78	4.95	5.09	5.21	5.31	5.40	6.08
10	3.58	3.83	4.01	4.15	4.27	4.37	4.45	4.53	4.59	5.06
12	3.43	3.65	3.80	3.93	4.04	4.13	4.20	4.26	4.32	4.73
15	3.29	3.48	3.62	3.74	3.82	3.90	3.97	4.02	4.07	4.42
20	3.16	3.33	3.46	3.55	3.63	3.80	3.76	3.80	3.85	4.15
24	3.09	3.26	3.38	3.47	3.54	3.61	3.66	3.70	3.74	4.04
30	3.03	3.19	3.30	3.39	3.46	3.52	3.57	3.61	3.65	3.90
40	2.97	3.12	3.23	3.31	3.38	3.43	3.48	3.51	3.55	3.79
60	2.92	3.06	3.16	3.24	3.30	3.34	3.39	3.42	3.46	3.69
120	2.86	2.99	3.09	3.16	3.22	3.27	3.31	3.34	3.37	3.58
∞	2.81	2.94	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.48

REFERENCES

Anderson, R. L. and Houseman, E. E. Table of orthogonal polynomial values extended to N = 104. Ames, Iowa: Iowa St. U. Res. Bulletin #297, 1942.

Bottenberg, R. A., and Ward, J. H. Applied multiple linear regression. Lackland Air Force Base, Texas: Personnel Research Laboratory PRL-TDR-63-6, 1963.

Cohen, J. Multiple regression as a general data-analytic system. Psychological Bulletin, 1963, 70, 426-443.

Cohen, J. and Cohen, P. Applied multiple regression/correlation analysis for the behavioral sciences. New York: Lawrence Erlbaum Associates, 1975.

Connett, W. A note on multiple comparisons. Multiple Linear Regression Viewpoints, 1971, 2, 23 -24.

Duncan, D. B. Multiple range and multiple F tests. Biometrics, 1955, 11, 1-42.

Dunn, U. J. Multiple comparisons among means. Journal of the American Statistical Association, 1961, 56: 52-64.

Dunnett, C. W. A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Association, 1955, 50: 1096-1121.

Dunnett, C. W. New tables for multiple comparisons with a control. Biometrics, 1964, 20, 482-491.

Edwards, A. L. Experimental design in psychological research. (4th Ed.) New York: Holt, Rinehart and Winston, 1972.

Harter, H. L. Tables of range and studentized range. Annals of Mathematical Statistics, 1960, 31, 1122-1147.

Hays, W. L. Statistics for the social sciences (2nd ed.) New York: Holt, Rinehart and Winston, 1973.

Jennings, E. Fixed effects analysis of variance by regression analysis. Multivariate Behavioral Research, 1967, 2, 95-108.

Kelly, F. J., Beggs, D. W., and McNeil, K. A. Multiple regression approach. Carbondale, Ill.: Southern Illinois University Press, 1969.

Kerlinger, F. N. and Pedhazur, E. J. Multiple regression in behavioral research. New York: Holt, Rinehart and Winston, 1974.

Keuls, M. The use of studentized range in connection with an analysis of variance. Euphytica, 1952, 1, 112-122.

Kimball, A. W. On dependent tests of significance in the analysis of variance. Annals of Mathematical Statistics, 1951, 22, 600-602.

Lindem, A. C. and Williams, J. D. A regression program for Tukey's test. Behavior Research Methods and Instrumentation, 1975, 7, 375.

McNeil, K. A. and Beggs, D. L. Directional hypotheses with the multiple linear regression approach. Multiple Linear Regression Viewpoints, 1971, 1, No. 3, 89-102.

McNeil, K. A., Kelly, F. J. and McNeil, J. T. Testing research hypotheses using multiple linear regression. Carbondale, Ill.: Southern Illinois University Press, 1975.

Mendenhall, W. The design and analysis of experiments. Belmont, Calif.: Wadsworth, 1968.

Miller, R. G. Simultaneous statistical inference. New York: McGraw-Hill, 1966.

Newman, D. The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation. Biometrika, 1939, 31, 20-30.

Newman, I., Deitchman, R., Burkholder, J., Sanders, R. and Ervin, L. Type VI error: Inconsistency between the statistical procedure and the research question. Multiple Linear Regression Viewpoints, 1976, 6, No. 4, 1-19.

Newman, I. and Fry, J. A note on multiple comparisons and a comment on shrinkage. Multiple Linear Regression Viewpoints, 1972, 2, No. 3, 36-39.

Ryan, T. A. Multiple comparisons in psychological research. Psychological Bulletin, 1959, 56, 26-47.

Ryan, T. A. The experiment as the unit for computing rates of error. Psychological Bulletin, 1962, 59, 301-305.

Scheffé, H. A method for judging all contrasts in the analysis of variance. Biometrika, 1953, 40, 87-104.

Scheffé, H. The analysis of variance. New York: John Wiley & Sons, 1959.

Tukey, J. W. The problem of multiple comparisons. Dittoed, Princeton University, 1953.

Ward, J. H. Synthesizing regression models--an aid to learning effective problem analysis. American Statistician, 1969, 23, 14-20.

Ward, J. H. and Jennings, E. E. Introduction to linear models. Englewood Cliffs, N. J.: Prentice-Hall, 1973.

Williams, J. D. A multiple regression approach to multiple comparisons for comparing several treatments with a control. Journal of Experimental Education, 1971, 39, 93-96.

Williams, J. D. Multiple comparisons in a regression approach. Psychological Reports, 1972, 30, 639-647.

Williams, J. D. Regression analysis in educational research. New York: MSS Information Corporation, 1974a.

Williams, J. D. A simplified regression formulation of Tukey's test. Journal of Experimental Education, 1974b, 42, No. 4, 80-82.

Williams, J. D. A regression formulation of Dunn's and Scheffé's tests. Multiple Linear Regression Viewpoints, 1975, 6, No. 1, 74-82.

Winer, B. J. Statistical principles in experimental design. (2nd Ed.) New York: McGraw-Hill, 1971.