MULTIPLE COMPARISONS IN A REGRESSION FRAMEWORK

John D. Williams
The University of North Dakota

In an analysis of variance framework, a great deal of effort has been expended in the past two decades with the multiple comparisons situation. Essentially, the concern has been to preserve the probability level in the experimental situation, and still make additional tests involving the means, in addition to the main effects test that is usually made in a one-way layout.

Within the analysis of variance framework, several tests for multiple comparison have been devised. Dunnett (1955, 1964) constructed a test applicable to the situation in which several experimental groups are to be compared to a control group. Duncan's (1955) test is useful to comparing each mean to every other mean. Dunn (1961) devised a test which would retain maximum power if a limited number of comparisons are of interest and are decided upon on an _a priori_ basis. A test which is useful on an _a posteriori_ basis is Scheffé's (1953) test. This test is amenable to data snooping, but has the drawback of losing power, as compared to the other methods.

Each of the previously mentioned tests require either additional tables or, in the case of Scheffé's test, a modification of the usual tables for the F test. On the other hand, these same tests can be achieved by using multiple regression as a problem solving technique. There is one logical extension here: _appropriate tables should be consulted_. This point will be elaborated on in more detail later.

An Illustrative Example

Suppose the following information were available on four random groups:

GROUP I	GROUP II	GROUP III	GROUP IV
9	8	13	15
8	7	10	12
6	8	12	10
3	6	11	17
4	6	14	11
$\overline{X}_1 = 6.0$	$\overline{X}_2 = 7.0$	$\overline{X}_3 = 12.0$	$\overline{X}_4 = 13.0$

The different types of multiple comparison procedures involve different hypotheses (i.e. restrictions). The various types of multiple comparison procedures to be considered in this paper are the following: Duncan's multiple range test, Dunn's "c" test, and Scheffé's test. Dunnett's test for several comparisons with a control has been treated elsewhere (Williams, in press).

Duncan's Multiple Range Test

For the data presented in the illustrative example, there are $\binom{4}{2}$, or 6, comparisons of interest (that is, all possible contrasts of pairs) for Duncan's multiple range 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 $\overline{X}_4$

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

The full model for the data in the example is:

$$Y = b_0U + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + E \qquad (1)$$

where

U = a unit vector

X_1 = 1 if the score is from a member of Group 1; and 0 otherwise

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

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

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

b_0 - b_4 are the regression coefficients determined by the least squares method

E_1 = the error involved in prediction

Restricted models in the regression framework are easily developed. For example, for the hypothesis $\overline{X}_1 = \overline{X}_2$, if the regression coefficients are equated in the full model ($b_1 = b_2 = b_6$), then the restricted model can be found:

$$Y = b_5U + b_6X_1 + b_6X_2 + b_6X_3 + b_6X_4 + E_2$$

$$Y = b_5U + b_6(X_1 + X_2) + b_7X_3 + b_8X_4 + E_2 \qquad (2)$$

Let V_1 =1 if the score is from a member of <u>either</u> X_1 or X_2; and 0 otherwise

Then equation (2) can be transformed:

$$Y = b_5U + b_6V_1 + b_7X_3 + b_8X_4 + E_2 \qquad (3)$$

Equation (3) in the <u>restricted model</u> for the hypothesis $\overline{X}_1 = \overline{X}_2$. Similar restricted models can be written for the remaining five comparisons. To make this more specific, Table 1 contains a useful formulation for this situation.

Table 1

A Regression Formulation of Duncan's Multiple Range Test

Y	U	X_1	X_2	X_3	X_4	V_1	V_2	V_3	V_4	V_5	V_6
9	1	1	0	0	0	1	1	1	0	0	0
8	1	1	0	0	0	1	1	1	0	0	0
6	1	1	0	0	0	1	1	1	0	0	0
3	1	1	0	0	0	1	1	1	0	0	0
4	1	1	0	0	0	1	1	1	0	0	0
8	1	0	1	0	0	1	0	0	1	1	0
7	1	0	1	0	0	1	0	0	1	1	0
8	1	0	1	0	0	1	0	0	1	1	0
6	1	0	1	0	0	1	0	0	1	1	0
6	1	0	1	0	0	1	0	0	1	1	0
13	1	0	0	1	0	0	1	0	1	0	1
10	1	0	0	1	0	0	1	0	1	0	1
12	1	0	0	1	0	0	1	0	1	0	1
11	1	0	0	1	0	0	1	0	1	0	1
14	1	0	0	1	0	0	1	0	1	0	1
15	1	0	0	0	1	0	0	1	0	1	1
12	1	0	0	0	1	0	0	1	0	1	1
10	1	0	0	0	1	0	0	1	0	1	1
17	1	0	0	0	1	0	0	1	0	1	1
11	1	0	0	0	1	0	0	1	0	1	1

To make the comparison of $\overline{X}_1$ to $\overline{X}_2$, the following equation can be used:

$$F' = \frac{\left(R^2_{FM} - R^2_{RM}\right)/1}{\left(1 - R^2_{FM}\right)/\,df_w} \qquad (4)$$

The R^2_{FM} is a term used for the square of the multiple correlation coefficient in the <u>full model</u>, and R^2_{RM} is a term for the square of the multiple correlation coefficient in the <u>restricted model</u>. The df_w term is equivalent to the degrees of freedom for within in an analysis of variance. situation; in the present situation, $df_w = 16$.

For the present comparison,

$R_{FM} = .84516$, and $R^2_{FM} = .71429$ ·

$R^2_{RM} = .83942$, and $R^2_{RM} = .70463$ ·

Using equation (4), $F' = .5414$

The focal question centers upon the evaluation of this number. One approach is simply to compare it to the F distribution with 1 and 16 degrees of freedom. Because the F distribution with 1 and k degrees of freedom is equal to t^2, it can be seen that, by using the F distribution in a straightforward manner, the evaluation has the same inherent problems as the usual t test. Also, in using Duncan's test, the experimenter <u>knows</u> he is going to make $\binom{n}{2}$ comparisons. Before answering directly the question concerning the evaluation of the outcome of $F' = .5414$, the other comparisons of interest are made.

A second comparison of interest in using Duncan's test is comparing $\overline{X}_1$ to $\overline{X}_3$:

The full model is:

$Y = b_0U + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + E_1$ (1)

With the restriction $b_1 = b_3 = b_{10}$,

$Y = b_9U + b_{10}X_1 + b_{11}X_2 + b_{10}X_3 + b_{12}X_4 + E_3$

$Y = b_9U + b_{10}(X_1 + X_3) + b_{11}X_2 + b_{12}X_4 + E_3$

Let $V_2 = 1$ if the score is from either a member of X_1 or X_3; 0 otherwise

Then

$Y = b_9U + b_{10}V_2 + b_{11}X_2 + b_{12}X_4 + E_3$ (5)

Equation (5) is the restricted model for the hypothesis $\bar{X}_1 = \bar{X}_3$; $R = .60564$ and $R^2 = .36680$. $F' = 19.4602$.

The additional comparisons were made by going through this procedure four more times.

For the comparison of $\bar{X}_1$ to $\bar{X}_4$, for the restricted model, $R = .49124$, and $R^2 = .24132$, with $F' = 26.4875$;

For the comparison of $\bar{X}_2$ to $\bar{X}_3$, for the restricted model, $R = .68773$, and $R^2 = .47297$, with $F' = 13.5144$;

For the comparison of $\bar{X}_2$ to $\bar{X}_4$, for the restricted model, $R = .60564$, and $R^2 = .36680$, with $F' = 19.4602$;

For the comparison of $\bar{X}_3$ to $\bar{X}_4$, for the restricted model, $R = .83943$, and $R^2 = .70464$, with $F' = .5414$.

Before interpreting these calculations, it is worthwhile to order the groups concerning the size of the means. The order from low to high is the same as the subscripts; that is, $\bar{X}_1$ is the lowest, $\bar{X}_2$ is the second lowest, $\bar{X}_3$ is next to highest, and $\bar{X}_4$ is highest.

To evaluate these calculations, in each case, the square root of the F' value is found. This number is then compared with the appropriate number from Duncan's tables (Duncan's tables can also be found in Edwards, 1968).

This is an important point: to make appropriate probability statements concerning the outcome of a series of comparisons, an appropriate table should be used. When making more than one comparison, the only times the F distribution could be directly used occur when the comparisons are orthogonal; even this concession to using the F distribution is sometimes disputed.

Table 2 summarizes the comparisons, using Duncan's multiple range test.

Table 2

Duncan's Multiple Range Test in a Regression Formulation

Comparison	F'	$\sqrt{F'} = t$	Region of Rejection at .05 level	Decision
$\bar{X}_1$ to $\bar{X}_2$	.5414	.735	$t \geqq 2.469$	retain H_0
$\bar{X}_1$ to $\bar{X}_3$	19.4602	4.411	$t \geqq 2.596$	reject H_0
$\bar{X}_1$ to $\bar{X}_4$	26.4875	5.147	$t \geqq 2.673$	reject H_0
$\bar{X}_2$ to $\bar{X}_3$	13.5144	3.680	$t \geqq 2.469$	reject H_0
$\bar{X}_2$ to $\bar{X}_4$	19.4602	4.411	$t \geqq 2.596$	reject H_0
$\bar{X}_3$ to $\bar{X}_4$	.5414	.736	$t \geqq 2.469$	retain H_0

If the F distribution had erroneously been used, the region of rejection would be $t = \sqrt{F_{1,16}} = \sqrt{4.49} = 2.12$. Thus, by using the tables for Duncan's multiple range test, it is less probable for the null hypothesis to be rejected. Of course, this is to be expected.

Dunn's "c" Test

Dunn's "c" test allows for a powerful multiple comparison method when the comparisons are planned beforehand and are few in number. Suppose the following four comparisons are of interest:

$\bar{X}_1$ to $\bar{X}_2$

$\bar{X}_3$ to $\bar{X}_4$

$\bar{X}_1$ to $\bar{X}_3$

$\bar{X}_1$ to $\frac{1}{3}\bar{X}_2 + \frac{1}{3}\bar{X}_3 + \frac{1}{3}\bar{X}_4$

The restricted models for the first three are identical to the same hypothesis in the previous section on Duncan's multiple range test, and the first three columns of Table 2 are relevant. For the final hypothesis, the restriction is

$$b_1 = \frac{1}{3}b_2 + \frac{1}{3}b_3 + \frac{1}{3}b_4 = \frac{1}{3}(b_2 + b_3 + b_4) = b_{14}$$

Since the full model is

$$Y = b_0U + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + E_1,$$

the restricted model is

$$Y = b_{13}U + b_{14}X_1 + \frac{1}{3}b_{14}(X_2 + X_3 + X_4) + E_4$$

$$Y = b_{13}U + \frac{b_{14}}{3}(3X_1 + X_2 + X_3 + X_4) + E_4$$

The full model, of course, is the same as equation (1). For the restricted model, $R = .56153$, with $R^2 = .31532$. Using equation (4), $F' = 22.3441$. As was the case for Duncan's multiple range test in Table 2, a table can be made for Dunn's "c" test. Before constructing the table, the t value is found by the transformation $t = F'$. These values for the first three comparisons are the same as in Table 2. For the last comparison, $t = \sqrt{22.3441} = 4.832$ Table 3 contains the comparisons listed in this section, using Dunn's "c" test as the multiple comparison technique.

Table 3

Dunn's "c" Test In A Regression Formulation

Comparison	F'	$\sqrt{F'} = t$	Region of Rejection at .05 level	Decision
$\bar{X}_1$ to $\bar{X}_2$	.5414	.735	$t \geq 2.818$	Retain
$\bar{X}_3$ to $\bar{X}_4$	.5414	.735	$t \geq 2.818$	Retain
$\bar{X}_1$ to $\bar{X}_3$	19.4602	4.411	$t \geq 2.818$	Reject
$\bar{X}_1$ to $\frac{1}{3}\bar{X}_2 + \frac{1}{3}\bar{X}_3 + \frac{1}{3}\bar{X}_4$	22.3441	4.832	$t \geq 2.818$	Reject

The critical values for this test are obtained from tables in Dunn's article. Again, these values are used rather than using the F distribution or the t distribution directly; the reason for using these tables is to preserve the apparent probability level.

Scheffé's Test

Scheffé's test will allow <u>any</u> comparison to be made, including any *a posteriori* comparisons that might be interesting to the researcher. This test does, however, have an accompanying loss of power. The same procedure for definition of full and restricted models is used (as was the case in the two previous sections of Duncan's multiple range test and Dunn's "c" test). The difference lies in the distribution to which the value found from equation (4) is to be compared; the correct distribution to be compared to is $(k-1)_{\alpha} F_{k-1,\, N-k}$.

While it is impossible to list all comparisons that might be considered (there are an infinite number of such comparisons), it should be pointed out that beyond the seven comparisons given in the two previous sections, comparisons such as:

$$\tfrac{1}{9}\bar{X}_1 + \tfrac{8}{9}\bar{X}_2 = \tfrac{3}{7}\bar{X}_3 + \tfrac{4}{7}\bar{X}_4$$

can be considered. The restrictions on the regression coefficients for such a comparison would be:

$$\tfrac{1}{9}b_1 + \tfrac{8}{9}b_2 = \tfrac{3}{7}b_3 + \tfrac{4}{7}b_4$$

A simpler expression of these restrictions is:

$$b_1 + 8b_2 = 3b_3 + 4b_4.$$

The same comparisons listed earlier are considered from the point of view of Scheffé's test, and the results can be found in Table 4.

Table 4

Scheffé's Test In a Regression Formulation

Comparison	F'	Region of Rejection at .05 level	Decision
$\bar{X}_1$ to $\bar{X}_2$	.5414	$F' \geq 9.72$	Retain
$\bar{X}_1$ to $\bar{X}_3$	19.4602	$F' \geq 9.72$	Reject
$\bar{X}_1$ to $\bar{X}_4$	26.4875	$F' \geq 9.72$	Reject
$\bar{X}_2$ to $\bar{X}_3$	13.5144	$F' \geq 9.72$	Reject
$\bar{X}_2$ to $\bar{X}_4$	19.4602	$F' \geq 9.72$	Reject
$\bar{X}_3$ to $\bar{X}_4$	.5414	$F' \geq 9.72$	Retain
$\bar{X}_1$ to $\frac{1}{3}\bar{X}_2 + \frac{1}{3}\bar{X}_3 + \frac{1}{3}\bar{X}_4$	22.3441	$F' \geq 9.72$	Reject

The region of rejection is defined by $(k-1)_{\alpha} F_{k-1,\ N-k}$ which is $3(3.24) = 9.72$.

SUMMARY

In using multiple regression as a problem solving technique, one problem that might arise is the overuse of a full model with several restricted models, without adjusting the probability level. Such an approach would violate the apparent probability level. This has long been a concern in statistics. Several multiple comparison procedures have been developed for different situations.

The intent of the present paper has been to extend some of the better known multiple comparison procedures to a multiple regression approach. The major change in the regression approach is to assess the result of multiple uses of a full model to a correct distribution, rather than a straight-forward usage of the F distribution.

References

Duncan, D. B. "Multiple Range and Multiple F-tests," Biometrics, 1955, 11: 1-42.

Dunn, O. 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 20: 482-491, 1964.

Edwards, A. L. Experimental Design in Psychological Research, 3rd Edition, New York: Holt, Rinehart and Winston, 1968.

Scheffé, H. "A Method for Judging all Contrasts in the Analysis of Variance," Biometrika, 1953, 40: 87-104.

Williams, J. D. "A Multiple Regression Approach to Multiple Comparisons For Comparing Several Treatments with a Control," Journal of Experimental Education (in press).