

## Multiple Linear Regression with Two Qualitative Independent Variables and Two-way ANOVA

With multiple regression it is possible to include both quantitative and qualitative independent variables. Before turning to regression models with both qualitative and quantitative variables, it will be useful to first focus on regression models with two qualitative independent variables. Note that the information presented here also holds for more than two qualitative independent variables, as will be illustrated elsewhere.

### *The Regression Equation*

Using fictional data, assume one wishes to know whether student scores in ITBS mathematics differ across teachers and student sex. Both sex and teacher variables are categorical, or qualitative, so both will require dummy variable coding to adequately represent the variables in the multiple regression equation. Since sex contains only two categories, male vs. female, one dummy variable will be needed. With the sample data provided in Table 1, the dummy variable representing sex is labeled MALE, with 1 = males and 0 = females.

The variable teacher contains three categories: Smith, Williams, and Miller. Two dummy variables will be needed to properly represent teacher in the regression equation. Students in Smith's class will serve as the reference category, so the two dummy variables entered are WILL (coded 1 for students in William's class, otherwise 0), and MILL (coded 1 for students in Miller's class and 0 otherwise). The sample regression equation, in which standardized test scores are the dependent variable and denoted  $Y$ , is

$$Y_i = b_0 + b_1\text{MALE}_{1i} + b_2\text{WILL}_{2i} + b_3\text{MILL}_{3i} + e_i, \quad (8)$$

In (8),  $b_0$  is the sample intercept;  $b_1$  is the sample partial regression coefficient for males controlling for the effects of teacher;  $b_2$  is the sample partial regression coefficient for Williams' class controlling for the effect of sex;  $b_3$  is partial coefficient for Miller's class controlling for sex; and  $e_i$  is the sample error term.

Since the variables are categorical, the partial regression coefficients represent the mean differences between the various categories controlling for the effect of the other independent variable. Thus,  $b_1$  is the estimated mean difference in standardized test scores between males and females controlling for teacher;  $b_2$  is the estimated mean difference between students in Williams' and Smith's class—since students in Smith's class are the reference group—controlling for sex differences; and  $b_3$  is the partial mean difference between Millers' and Smith's classes, where sex is partialled or taken into account.

The population regression equation for this model is

$$Y_i = \beta_0 + \beta_1\text{MALE}_{1i} + \beta_2\text{WILL}_{2i} + \beta_3\text{MILL}_{3i} + \varepsilon_i, \quad (9)$$

where,

$Y_i$  is  $i^{\text{th}}$  student's achievement score;

$\beta_1$  is the population partial regression coefficient expressing the difference between males and females given (or controlling for) teacher;

$\beta_2$  is the population partial regression coefficient indicating the mean difference between students in Williams' and Smith's classes, controlling for sex;

$\beta_3$  is the partial coefficient representing the mean difference between Miller's and Smith's classes, controlling for sex;

$\beta_0$  is the population intercept for the equation; and  $\varepsilon_i$  is the a random error.

Table 1  
*Sample Data for ITBS, Sex, and Teacher*

ITBS Math	Sex	Teacher	MALE	WILL	MILL
51	M	Smith	1	0	0
53	M	Smith	1	0	0
55	M	Smith	1	0	0
56	M	Williams	1	1	0
55	M	Williams	1	1	0
57	M	Williams	1	1	0
61	M	Miller	1	0	1
59	M	Miller	1	0	1
60	M	Miller	1	0	1
58	M	Miller	1	0	1
49	F	Smith	0	0	0
48	F	Smith	0	0	0
47	F	Smith	0	0	0
50	F	Smith	0	0	0
49	F	Williams	0	1	0
51	F	Williams	0	1	0
52	F	Williams	0	1	0
59	F	Miller	0	0	1
60	F	Miller	0	0	1
59	F	Miller	0	0	1
58	F	Miller	0	0	1

Note. F = female, M = male, WILL = William's class, and MILL = Miller's class.

Referring back to equation (1), the sample regression model, note that  $b_0$  represents the mean ITBS score for female students in Smith's class. Why? Recall that for the dummy variable MALE, 0 represents females. Also, for the teacher dummy variables, WILL and MILL, the reference or comparison group is Smith's class. Taken together, the overall comparison group is female students with Smith. Further discussion of comparisons and interpretation of regression coefficients will be presented below.

### ***The Prediction Equation and Residuals***

The prediction equation for the example data is:

$$Y' = b_0 + b_1\text{MALE}_{1i} + b_2\text{WILL}_{2i} + b_3\text{MILL}_{3i}. \quad (10)$$

OLS estimates for this model are:

$$Y' = 49.05 + 3.20(\text{MALE}_{1i}) + 2.67(\text{WILL}_{2i}) + 8.59(\text{MILL}_{3i}).$$

Residuals are obtained in a manner identical to that described earlier. Namely, one obtains the predicted value  $Y'$  and subtracts this value from the observed  $Y$ , i.e.,

$$e_i = Y - Y'.$$

Consider, for instance, the residual for the first individual in Table 1. Note that this individual is a male in Smith's class, so only the sex dummy variable, MALE, will have a nonzero value. The predicted ITBS score for this individual is:

$$\begin{aligned} Y' &= 49.05 + 3.20(\text{MALE}_{1i}) + 2.67(\text{WILL}_{2i}) + 8.59(\text{MILL}_{3i}), \\ &= 49.05 + 3.20(1) + 2.67(0) + 8.59(0), \\ &= 49.05 + 3.20(1), \\ &= 49.05 + 3.20, \\ &= 52.25. \end{aligned}$$

This individual's observed ITBS value is 51, so the residual is:

$$\begin{aligned} e_1 &= Y - Y', \\ &= 51 - 52.25 = -1.25. \end{aligned}$$

The residual indicates that this person's score is over-predicted; that is, the observed score is less than the predicted score.

As previously noted, the goal of OLS is to obtain estimates for the regression coefficients that will provide the smallest possible residuals for all observations in the sample, and when certain assumptions are met, OLS estimates do provide the smallest possible residuals (for proof, see the Pedhazur text). In short, OLS attempts to find the regression line that passes through all observations and that provides the smallest set of squared residuals,  $\sum e_i^2$ .

### ***Regression Coefficient Interpretation***

As previously mentioned, the intercept,  $b_0$ , represents the mean ITBS score for the omitted categories, which, for the current example, are females in Smith's class. The sex coefficient,  $b_1$ , indicates the estimated mean difference between males and females, controlling for teacher. Since there are no interactions (to be described later) between the two independent variables, the estimated difference between males and females will be the same for all categories of teacher. That is, the estimated mean difference between males and females is  $b_1 = 3.20$  points, and this 3.20 point spread is the expected difference between males and females for students in each of the three classes.

This difference can be illustrated for the three classes by examining the predicted means for males and females for each class. For example, the predicted mean for Smith's females is

$$Y' = 49.05 + 3.20(0) + 2.67(0) + 8.59(0) = 49.05,$$

and for Smith's males the predicted mean is

$$Y' = 49.05 + 3.20(1) + 2.67(0) + 8.59(0) = 52.25.$$

Note that the difference between Smith's males and females is  $52.25 - 49.05 = 3.20$ . Similarly, the predicted mean for William's females is

$$Y' = 49.05 + 3.20(0) + 2.67(1) + 8.59(0) = 51.72,$$

and for William's males

$$Y' = 49.05 + 3.20(1) + 2.67(1) + 8.59(0) = 54.92.$$

Again, the predicted difference is  $54.92 - 51.72 = 3.20$ . Finally, the predicted means for Miller's females and Miller's males, respectively are

$$Y' = 49.05 + 3.20(0) + 2.67(0) + 8.59(1) = 57.64, \text{ and}$$

$$Y' = 49.05 + 3.20(1) + 2.67(0) + 8.59(1) = 60.84,$$

which results in a mean difference of  $60.84 - 57.64 = 3.20$ .

Thus, as stated above, the estimated difference between males and females,  $b_1$ , is the same across all three classes. Note, also, that the estimated the mean difference among the classes are the same for both males and females. To illustrate, the mean difference between Miller's class and Smith's class is  $b_3$ , or 8.59 points. This 8.59 point difference is the same both for males and for females. For example, the predicted mean ITBS score for Smith's females is

$$Y' = 49.05 + 3.20(0) + 2.67(0) + 8.59(0) = 49.05,$$

and for Miller's females the predicted mean is

$$Y' = 49.05 + 3.20(0) + 2.67(0) + 8.59(1) = 57.64.$$

The difference between these two means is  $57.64 - 49.05 = 8.59$  or  $b_3$ . Similarly, the difference between the Smith's and Miller's males will also be 8.59 points as indicated by the predicted means for Smith's and Miller's classes, respectively:

$$Y' = 49.05 + 3.20(1) + 2.67(0) + 8.59(0) = 52.25, \text{ and}$$

$$Y' = 49.05 + 3.20(1) + 2.67(0) + 8.59(1) = 60.84,$$

where  $60.84 - 52.25 = b_3 = 8.59$ .

If one were to similarly calculate the predicted means for Williams' and Smith's classes, the difference, whether for males or females, would be  $b_2$  or 2.67 points. Should one wish to compare Williams' and Miller's classes, the difference, as explained earlier when discussing regression with one qualitative predictor, is

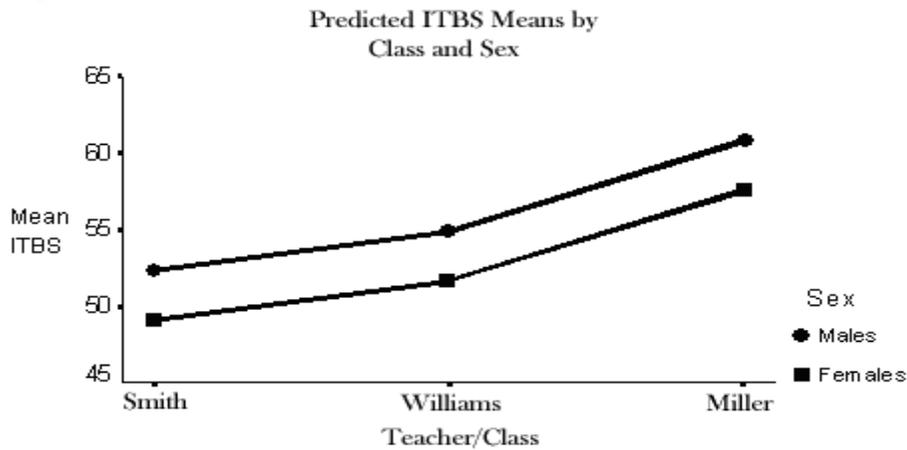
$$\text{Williams' class vs. Miller's class} = b_2 - b_3 = 2.67 - 8.59 = -5.92.$$

The calculated difference between Williams' and Miller's classes of -5.92 is the same whether one is examining males or females.

The above calculations show that the statistical effects due to teacher are the same whether one is considering males or females; similarly, the differences between males and females is the same whether one is examining differences for Williams', Miller's or Smith's classes. The constant differences in sex among the various classes, and in class between the sexes, can be graphically displayed, as is done in Figure 1 below.

As Figure 1 illustrates, the average difference between males and females is the same all classes examined. Also, the predicted difference among the classes is the same for both males and females for each respective pairwise comparison, i.e., the difference by sex is the same in Miller's class as in Smith's class.

Figure 2



The regression coefficients for dummy variables in multiple regression essentially retain the interpretations first discussed for the situation of a single independent variable. The primary difference is that the coefficients represent partial effects in which the other independent variables are now taken into account when estimating the mean differences among groups. To illustrate how sex is taken into account, consider the following model:

$$Y' = b_0 + b_1 \text{MALE}_{i_i}$$

Note that in the multiple regression model in which teacher was factored into the equation the estimated difference between the sexes was 3.20 points. When teacher is not controlled, OLS produces the following results:

$$Y' = 52.91 + 3.59(\text{MALE}_{i_i})$$

The estimated difference, ignoring differences in teachers, between males and females is 3.59 points, but once teacher is statistically modeled, the difference is estimated to be 3.20 points. This difference in estimates suggests that some of the difference attributed to sex discrepancies may be due, instead, to teacher differences in ITBS performances.

### ***Overall Model Fit and Statistical Inference***

Previous discussions of model fit (or the lack thereof) continue to hold here. As before, the initial hypothesis tested for tenability should be the overall fit hypothesis:  $H_0: R^2 = 0.00$ . The usually formulas for calculating the overall F ratio apply, such as

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

where

k = number of independent variables (or vectors) in the model; and

n = sample size (or number of observations in sample).

The observed  $R^2$  for the current example .8754. Using the  $R^2$  formula, the calculated overall F ratio is

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} = \frac{.8753/3}{(1 - .8753)/(21 - 3 - 1)} = 39.80.$$

With  $df_1 = k = 3$  and  $df_2 = n - k - 1 = 21 - 3 - 1 = 17$ , the .05 level critical F value is

$$.05F_{3,17} = 3.20.$$

Since 39.80 is larger than the critical F of 3.20,  $H_0$  is rejected and one may conclude that the model of ITBS scores using sex and teacher explains statistically more variability than would be expected by chance alone. One may also use p-values to test the null. The corresponding p-value of an F of 39.80 with degrees of freedom of 3 and 17 is .0000 which is smaller than .05.

### ***Global Effects, $\Delta R^2$ , and the Partial F Test of $\Delta R^2$***

To test whether a given categorical independent variable with *more than two* categories makes a statistically significant contribution to the model of Y, one must assess the global effect of that independent variable. The global effect is defined as the  $\Delta R^2(X_k)$  associated that the categorical independent variable in question. For example, to determine whether teacher makes a statistically significant contribution to the explained variance or the reduction in error variance of ITBS scores, one must test the null hypothesis

$$H_0: \Delta R^2(\text{teacher}) = \Delta R^2(X_2, X_3) = 0.00.$$

This null indicates that the combination of dummies representing teacher,  $X_2$  and  $X_3$ , do not make an increase in the overall model fit as measured by the model  $R^2$ .

The earlier discussion of  $\Delta R^2(X_k)$  illustrated how one calculates and tests the partial increment in  $R^2$ . To do so requires that one first estimate the regression model with all relevant independent variables included except for the variable to be tested—this will be the reduced model. Next, one enters the independent variable to be tested and re-estimates the model—this is the full model. To illustrate, the reduced model contains all variables except teacher. In this example there is only one other independent variable, sex, so the reduced model is

$$Y' = b_0 + b_1 \text{MALE}_{1i}.$$

The resulting statistics from the reduced model needed to perform the partial F test are:  $R_r^2 = .1652$  and  $df_{2r} = 19$ .

The full model contains all independent variables, including teacher:

$$Y' = b_0 + b_1 \text{MALE}_{1i} + b_2 \text{WILL}_{2i} + b_3 \text{MILL}_{3i}.$$

The full model statistics needed are:  $R_f^2 = .8754$  and  $df_{2f} = 17$ .

The increment in the model  $R^2$  due to adding teacher to the model, denoted  $\Delta R^2(X_2, X_3)$  or  $\Delta R^2(\text{teacher})$ , is

$$\Delta R^2(X_2, X_3) = R_f^2 - R_r^2 = .8754 - .1652 = .7102.$$

Calculation of the partial F test follows:

$$F = \frac{\Delta R^2(X_k)/(df_{2r} - df_{2f})}{(1 - R_r^2)/df_{2r}}$$

$$= \frac{.7102/(19 - 17)}{(1 - .1652)/17} = \frac{0.3551}{(0.1246)/17} = \frac{.3551}{0.0073} = 48.64.$$

Degrees of freedom for this test are

$$df_1 = df_{2r} - df_{2f} = 19 - 17 = 2, \text{ and}$$

$$df_2 = df_{2f} = 17.$$

The critical value at the 5% significance level is  $.05F_{2,17} = 3.59$ . Since 48.64 is larger than 3.59, the null is rejected and one may conclude that inclusion of teacher statistically reduces variance error in ITBS mathematics scores.

It is important that the global effect and tests of statistical significance for categorical independent variables with more than two categories be performed and reported. Although not likely, it is possible that while the global effect of a categorical variable is not significant, specific pairwise comparisons within the variable may be statistically significant. For example, assuming that the partial F test for teacher indicated the variable did not contribute to the overall fit of ITBS scores, it is still possible that the pairwise difference between, say, students in Smith's class and students in Miller's class could be statistically significant. Thus, the cautious researcher will plan ahead with specific, a priori comparisons to be performed. In addition, some control procedure for the inflation of the familywise error rate may be needed, such as Bonferroni (for planned comparisons) or Scheffé (for planned or post hoc). The use of these procedures will be illustrated below.

While one may use the partial F test for all variables (and blocks or sets of variables), it is not necessary for dichotomous and quantitative variables since the t-ratio,  $b_k/se_{bk}$ , will be equivalent to the partial F test for these variables. Variables that consume only one vector will have partial F numerator degrees of freedom,  $df_1$ , equal to one, and as has already been shown, F ratios with  $df_1$  equal to one are equal to squared t ratios, i.e.,

$$t^2 = F_{1,\infty}.$$

The use of partial F tests for more than one variable, i.e., sets or blocks of variables, will be illustrated below.

### ***Inferential Procedures for Regression Coefficients***

Following the rejection of the overall model null  $H_0: R^2 = 0.00$ , one should next test the partial effect of each distinct independent variable. This has just been illustrated for teacher using  $\Delta R^2(X_2, X_3)$  and the partial F test. Since the remaining variable, sex, is dichotomous, the simple t-ratio will suffice for testing  $H_0: \beta_1 = 0.00$ . To illustrate the redundancy of the partial F test to the t-ratio for one vector variables, those with only 1 degree of freedom, both the t and partial F will be used to test the effect of sex.

The partial regression coefficient and standard error, both obtained from OLS, for sex are  $b_1 = 3.205$  and  $se_{b_1} = 0.758$ . The corresponding t ratio is  $3.20/0.758 = 4.228$ . The two-tailed critical t,  $_{\alpha/2}t_{df_2} = .05/2t_{17} = 2.11$ , where  $df = n - k - 1$ . Since t is larger than  $t_{crit}$ ,  $H_0$  is rejected, and one may conclude that ITBS scores differ by sex, or, equivalently, that sex contributes to a reduction in error variance associated with ITBS scores—i.e., that sex contributes to the model of ITBS scores.

If one wishes to use the partial F test for sex, the appropriate null would be  $H_0: \Delta R^2(X_1) = 0.00$ . Note that this null and  $H_0: \beta_1 = 0.00$  are equivalent since there is only 1 degree of freedom for this variable. The reduced model contains only teacher, i.e.,

$$Y = b_0 + b_2WILL_{2i} + b_3MILL_{3i} + e_i,$$

with  $R^2_f = .7444$  and  $df_{2r} = n - k_r - 1 = 21 - 2 - 1 = 18$ . The full model is

$$Y = b_0 + b_1\text{MALE}_{1i} + b_2\text{WILL}_{2i} + b_3\text{MILL}_{3i} + e_i,$$

with  $R_f^2 = .8754$  and  $df_{2f} = n - k_f - 1 = 21 - 3 - 1 = 17$ .

The semi-partial effect of sex is

$$\Delta R^2(X_1) = R_f^2 - R_r^2 = .8754 - .7444 = .131.$$

The partial F statistic for sex is

$$F = \frac{\Delta R^2(X_k)/(df_{2r} - df_{2f})}{(1 - R_f^2)/df_{2f}}$$

$$= \frac{.131/(18 - 17)}{(1 - .8754)/17} = \frac{.131}{(0.1246)/17} = \frac{.131}{0.0073} = 17.945.$$

with

$$df_1 = df_{2r} - df_{2f} = 18 - 17 = 1, \text{ and}$$

$$df_2 = df_{2f} = 17.$$

The 5% critical value  $.05F_{1,17} = 4.45$ . Since 17.95 larger than 4.45, the null is rejected, and one may conclude that the inclusion within the model of sex statistically reduces variance error in ITBS scores.

As a comparison, the t ratio for sex is 4.228 and the F is 17.95. Recall that the relationship between t and F when  $df_1 = 1$  is

$$t^2 = F_{1,\infty};$$

and  $4.228^2 = 17.86$ , which is within rounding error of 17.95.

### ***Pairwise Comparisons Among Categories and Control of the Familywise Type I Error Rate***

In the current regression example, there are two qualitative independent variables. For sex, there are only two categories, so at most there is one comparison, but for teacher, there are three categories, so there are  $C = G(G-1)/2 = 3(3-1)/2 = 3$  possible pairwise comparisons among the classes. The variable teacher, then, serves as a family of possible comparisons. As a result, one must be concerned with the possible inflation of the Type I error rate associated with this family of comparisons. Either the Bonferroni or Scheffé procedure may be used to control  $\alpha_{fw}$  at an appropriate level. For illustrative purposes, assume that post hoc comparisons are being performed, so the Scheffé method will be used.

The first step is to obtain the estimated mean differences among teachers in terms of ITBS scores. In the sample regression equation, Smith's class is the comparison group, so  $b_2$  represents the mean difference between Williams' and Smith's classes, controlling for sex, and  $b_3$  the difference between Miller's and Smith's classes, again, controlling for sex. The standard errors for these differences are, respectively,  $se_{b_1}$  and  $se_{b_2}$ . The last pairwise comparison is Williams' class vs. Miller's class. In using the computer approach to calculating this difference (see discussion of regression with one qualitative IV), Miller's class will serve as the reference group and the revised sample regression equation is

$$Y = b_0 + b_1\text{MALE}_{1i} + b_2\text{SMITH}_{2i} + b_3\text{WILL}_{3i} + e_i. \tag{11}$$

In equation (11),  $b_3$  represents the comparison between Williams' class and Miller's class. The estimated mean difference, controlling for sex, is  $b_3 = -5.917$  with a standard error of 0.935.

Table 4 provides a summary of the obtained comparisons.

Table 4  
*Comparisons of ITBS Scores Among Teachers Controlling for Student Sex*

Contrast	Estimated Mean Difference	Standard Error of Difference	t-ratio
Williams vs. Smith	2.676	0.965	2.77
Miller vs. Smith	8.592	0.898	9.57
Williams vs. Miller	-5.917	0.935	-6.33

To assess statistical significance for the comparisons requires appropriate critical values. Since these comparisons are, arbitrarily, being considered post hoc, a Scheffé critical t value will be used to test

$$H_0: \mu_i = \mu_j$$

where  $i$  represents one group and  $j$  another group, such as  $i = \text{Williams}$  and  $j = \text{Miller}$ .

To find Scheffé critical t-ratios, the appropriate critical F value must first be obtained. If one wishes to set  $\alpha_{ew}$  (or similarly, in this case,  $\alpha_{fw}$ ) at .05, then the Scheffé critical F is

$$\begin{aligned} \text{Scheffé } F_{\text{crit}} &= (J - 1)(F_{\alpha_{ew}, J-1, n-k-1}), \\ &= (3 - 1)(F_{.05, 3-1, 21-3-1}), \\ &= (2)(F_{.05, 2, 21-3-1}), \\ &= (2)(F_{.05, 2, 21-3-1}), \\ &= (2)(3.59) = 7.18. \end{aligned}$$

where  $J = \text{number of groups}$ . The comparable Scheffé critical t is

$$\begin{aligned} \text{Scheffé } t_{\text{crit}} &= \sqrt{(J - 1)(F_{\alpha_{ew}, J-1, n-k-1})} = \sqrt{\text{Scheffé } F_{\text{crit}}} \\ &= \sqrt{7.18} = \pm 2.6796. \end{aligned}$$

To complete the pairwise comparisons, one must simply calculate the t ratio for each comparison by dividing the estimated mean difference by its standard error. The resulting t ratio should then be compared to the Scheffé  $t_{\text{crit}}$  to determine whether a statistically significant difference exists. For example, the calculated t ratio for the Williams vs. Smith comparison, based upon the information provided in Table 4, is  $-5.917/0.935 = -6.33$ . The usual decision rule is applied, namely:

**If  $t \leq \text{Scheffé } -t_{\text{crit}}$  or  $t \geq \text{Scheffé } t_{\text{crit}}$  then reject  $H_0$ , else FTR  $H_0$ .**

Since -6.33 is less than -2.6796, the null hypothesis of no difference is rejected and one may conclude that mean levels of ITBS scores differ between Williams' class and Smith's class, with Williams' students obtaining the higher mean score. Similar testing is done for the remaining comparisons.

#### **Confidence Intervals (CI)**

CIs may, and should, be constructed for each comparison. When a multiple comparison procedure is used, CI for those comparisons performed under the multiple comparison procedure should be based upon the corrected or adjusted critical values used to control for inflation of  $\alpha_{ew}$ . In the current example,

for instance, CIs for contrasts among the teachers' classes will be computed using the Scheffé adjusted critical t-ratios, and the CI for the sex difference will be based upon a non-corrected or traditional critical t since multiple comparison procedures were not used, or needed, for this variable.

To illustrate, the .95CI for the sex difference is

$$.95CI: b_1 \pm t_{(\alpha/2, df)}se_{b1}.$$

Calculated, the sex .95CI is 1.61 and 4.81. The .95CI for the teacher comparisons are

$$.95CI: b_2 \pm (\text{Scheffé } t_{crit})se_{b2}, \text{ and}$$

$$.95CI: b_3 \pm (\text{Scheffé } t_{crit})se_{b3},$$

$$.95CI: b_2-b_3 \pm (\text{Scheffé } t_{crit})se_{b2-b3}.$$

Remember that one may calculate the difference between Williams' class and Miller's class,  $b_2-b_3$ , and its standard error either by hand or by computer. The .95CI for the William's class vs. Smith's class comparison, for example, is

$$.95CI: 2.676 \pm (2.6796)(0.965),$$

$$.95CI: 2.676 \pm 2.586,$$

$$.95CI: 5.262, 0.090.$$

All CIs are calculated and reported in the following section.

### ***Reporting Multiple Regression with Two Qualitative IVs***

Sample results for the ITBS data are provided below. For variety, and to test your perceptive abilities, the results were obtained from STATA, not SPSS, this time.

```
. regress itbs male will mill
```

Source	SS	df	MS	Number of obs =	21
Model	357.982061	3	119.327354	F( 3, 17) =	39.80
Residual	50.9703196	17	2.9982541	Prob > F =	0.0000
				R-squared =	0.8754
				Adj R-squared =	0.8534
Total	408.952381	20	20.447619	Root MSE =	1.7315

itbs	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
male	3.205479	.7582926	4.227	0.001	1.605622 4.805337
will	2.675799	.9648651	2.773	0.013	.6401116 4.711487
mill	8.592466	.8977959	9.571	0.000	6.698282 10.48665
_cons	49.05479	.730709	67.133	0.000	47.51313 50.59646

Reporting results may take several forms. Most common is a tabular display, although for models with few IVs, reporting results within the text of your manuscript may be feasible. However, the tabular format is preferred and will be illustrated.

**Clarification Note to Instructor:** Multiple comparisons reported below in Table 3 disagree with Scheffé results reported in SPSS and on Excel comparison spreadsheet found here:

[http://www.bwgriffin.com/gsu/courses/edur8131/content/MultipleComparisons-BonferroniandScheffe\\_revised.xls](http://www.bwgriffin.com/gsu/courses/edur8131/content/MultipleComparisons-BonferroniandScheffe_revised.xls)

The difference results from which group means are used to calculate comparisons. SPSS uses observed, unadjusted means across both sexes (i.e., estimated marginal means are not used for comparisons) when both teacher and sex are treated as factors:

Teacher	M	N
2 Williams	53.333	6
3 Miller	59.25	8
1 Smith	50.4286	7

With these mean differences

Contrast	Estimated Mean Difference	Standard Error of Difference	Scheffé Adjusted .95CI
W vs S (2 vs 1)	2.9048	.9633	0.322, 5.486
M vs S (3 vs 1)	8.8214	.8961	6.419, 11.22
W vs M (2 vs 3)	-5.9167	.9351	-8.423, -3.41

However, if sex is entered as a covariate, then results are same as regression with adjusted marginal means. Below are the adjusted marginal means controlling for sex (set at its mean of .4762 rather than 0 or 1). Due to unequal sample size for males and females, the following adjusted means result for each class:

Teacher	Adjusted M	N
2 Williams	<b>53.2572</b>	6
3 Miller	<b>59.1732</b>	8
1 Smith	<b>50.5812</b>	7

If the adjusted means, or estimated marginal means in SPSS, are used, then multiple comparisons that result are those reported in Table 3 in the example below. These results were confirmed using the Excel spreadsheet. Following criteria used for Excel spreadsheet:

DF1 = 2, DF2 = 17, Alpha = .05, Groups = 3, MSE = 2.998, means and n as noted above.

*Table 1*  
*Descriptive Statistics and Correlations Among ITBS Scores, Sex, and Teacher*

Variable	Correlations			
	ITBS	Male	Williams	Miller
ITBS	---			
Male	.406	---		
Williams	-.184	.030	---	
Miller	.823*	.037	-.496*	---
Mean	54.619	0.476	0.256	0.381
SD	4.522	0.512	0.463	0.498

*Note:* Male (male = 1, female = 0) is a dummy variable, as are William (=1, others = 0) and Miller (=1, others = 0); n = 21.

\*p < .05.

*Table 2*  
*Regression of ITBS Scores on Sex and Teacher*

Variable	b	se	$\Delta R^2$	95% CI	F	t
Male	3.205	0.758	.131	1.61, 4.81		4.23*
Teacher			.710		48.44*	
Williams	2.676	0.965		0.64, 4.71		2.77*
Miller	8.592	0.898		6.70, 10.49		9.57*
Intercept	49.055	0.731		47.51, 50.60		67.13*

*Note:*  $R^2 = .875$ , adj.  $R^2 = .853$ ,  $F_{3,17} = 39.80^*$ ,  $MSE = 2.998$ , n = 21.  $\Delta R^2$  represents the semi-partial correlation or the increment in  $R^2$  due to adding the respective variable. Male (male = 1, female = 0) is a dummy variable, as are William (= 1, others = 0) and Miller (= 1, others = 0).

\*p < .05.

*Table 3*  
*Comparisons of Adjusted Mean ITBS Scores Among Teachers*

Contrast	Estimated Mean Difference	Standard Error of Difference	Scheffé Adjusted .95CI
Williams vs. Smith	2.676*	0.965	0.09, 5.26
Miller vs. Smith	8.592*	0.898	6.19, 11.00
Williams vs. Miller	-5.917*	0.935	-8.42, -3.41

\*p < .05, where p-values are adjusted using the Scheffé method.

Regression results show that both student sex and teachers are statistically related to students' ITBS mathematics scores, and predictors account for, statistically, a large amount of variation in these scores. Once teacher is taken into account, males average about 3.20 points higher than females. Among the three teachers considered, Miller's students' score, on average, higher than both Williams' students and Smith's students by 5.92 and 8.59 points, respectively, and these differences are statistically significant at the .05 level. The average mean ITBS score for Williams' students, controlling for sex, is estimated to be about 2.68 points higher than Smith's students, and this difference is also statistically significant.

### Exercises

Assume all comparisons were planned in advanced.

(1) According to the leadership literature, there are a number of different leadership styles. Listed below are scores obtained from an instrument designed to measure a particular leadership style, which will be referred to as style X. Of interest is whether X differs by school district type in terms of urbanity, and by sex. A stratified random sample of school principals were selected from three district types (mostly urban, mostly suburban, and mostly rural).

The scores on style X range from 100 to 0. The closer the score to 100, the more the respondent conforms to style X, while the closer the score to 0, the less the respondent conforms to style X.

Is there any evidence that X differs among the three district types, or by sex?

Sex	District Type	Style X
m	urban	85
m	urban	98
m	urban	75
f	urban	63
m	urban	91
f	urban	49
f	urban	62
f	suburban	49
f	suburban	48
m	suburban	56
m	suburban	78
f	suburban	35
m	suburban	50
m	rural	33
m	rural	95
f	rural	26
f	rural	11
f	rural	33
m	rural	25
m	rural	65

(2) A researcher is interested in learning whether frequency of reading at home to elementary-aged children produces differential effects on reading achievement. After obtaining information from a randomly selected sample of parents about this behavior, the following classifications and standardized achievement scores were recorded. (Note: frequency classifications as follows: a = less than once per month, b = once to three times per month, c = more than three times per month.) In addition to reading frequency, information regarding the family's status concerning whether or not the family's child receives either free or reduced lunch is recorded as a proxy for SES.

SES	Freq. of Reading	Achievement
fr	a	48
fr	a	37
no	a	47
no	a	65
no	b	57
fr	b	39
fr	b	49
no	b	45
no	c	61
no	c	55
fr	c	51
fr	c	30

Note. FR indicates free or reduced lunch received, NO indicates otherwise.

Is frequency of reading at home related to student reading achievement once SES is taken into account?

(3) An administrator wishes to know whether student behavioral problems can be linked to student performance. If students were suspended or reprimanded more than once, they are classified as having behavioral problems. In addition, each student's SES is known, and should be taken into account. The administrator randomly selects 13 students and collects the appropriate data.

Student	GPA	Student SES	Behavioral Problems
Bill	3.33	h	n
Bob	1.79	l	y
Stewart	2.21	m	n
Linda	3.54	h	y
Lisa	2.89	m	n
Ann	2.54	m	n
Fred	2.66	h	y
Carter	1.10	l	y
Bill	3.10	h	n
Sue	2.10	l	y
Kara	2.07	l	y
Loser	2.31	m	n
Kathy	3.67	h	n

*Two-way ANOVA for Sample Data*

ANOVA VARIABLES=itbs BY teacher(1 3) sex(0 1) /MAXORDERS NONE  
/METHOD UNIQUE /FORMAT LABELS .

\* \* \* A N A L Y S I S O F V A R I A N C E \* \* \*

ITBS  
by TEACHER  
SEX

UNIQUE sums of squares  
All effects entered simultaneously

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	357.982	3	119.327	39.799	.000
TEACHER	290.439	2	145.219	48.435	.000
SEX	53.577	1	53.577	17.869	.001
Explained	357.982	3	119.327	39.799	.000
Residual	50.970	17	2.998		
Total	408.952	20	20.448		

21 cases were processed.  
0 cases (.0 pct) were missing.

## Results for Exercises

(1) Results for leadership style analysis.

*Table 1a*  
*Descriptive Statistics for Leadership Style, District Type, and Sex*

Variable	Correlations			
	Style	Urban	Suburban	Male
Style	---			
Urban	.55*	---		
Suburban	-.10	-.48*	---	
Male	.54*	.03	-.07	---
Mean	56.35	.350	.300	.550
SD	25.07	.489	.470	.510

Note: Male is a dummy variable (male = 1, female = 0), as are Urban (1, 0 = other) and Suburban (1, 0 = other); n = 20.

*Table 1b*  
*Regression of Style on Sex and District Type*

Variable	b	se	$\Delta R^2$	95% CI	F	t
Male	26.29	7.53	.29	10.32, 42.26		3.49*
District Type			.33		7.14*	
Urban	33.57*	8.94		14.62, 52.52		3.76*
Suburban	13.40	9.32		-6.36, 33.16		1.44
Intercept	26.12	7.65		9.91, 42.33		3.42*

Note:  $R^2 = .625$ , adj.  $R^2 = .555$ ,  $F_{3,16} = 8.90$ ,  $MSE = 279.70$ , n = 20.  $\Delta R^2$  represents the semi-partial correlation or the increment in  $R^2$  due to adding the respective variable. Male is a dummy variable (male = 1, female = 0), as are Urban (1, 0 = other) and Suburban (1, 0 = other).

\*p < .05.

*Table 1c*  
*Comparisons of Style Scores Among Urban, Suburban, and Rural Principals*

Contrast	Estimated Mean Difference	Standard Error of Difference	Bonferroni Adjusted 95% CI
Urban vs. Rural	33.57*	8.94	9.74, 57.40
Suburban vs. Rural	13.40	9.32	-11.44, 38.24
Urban vs. Suburban	20.17	9.32	-4.67, 45.01

\*p < .05, where p-values are adjusted using the Bonferroni method.

[Note, Bonferroni CI taken from Excel Spreadsheet is incorrect so must calculate CI using tabled values for Bonferroni comparisons. Use male = .55 in regression equation to obtain estimated means for each district. ]

Both sex and district type are statistically related to leadership style. Once district type is taken into account, males average about 26 points higher than females. Among the three district types considered, principals in urban settings have a statistically higher score on style than do principals in rural districts, but not statistically higher than principals in suburban districts.

(2) Results for reading frequency.

Table 2a  
Descriptive Statistics for Achievement, SES, and Reading Frequency

Variable	Correlations			
	Achievement	B	C	SES
Achievement	---			
B = 1 to 3 per month	-.09	---		
C = more than 3 per month	.04	-.50	---	
SES	-.65*	.00	.00	---
Mean	48.66	.333	.333	.500
SD	10.129	.492	.492	.522

Note: SES is a dummy variable (free/reduced lunch = 1, otherwise = 0), as are B (1, 0 = other) and C (1, 0 = other); n = 12.

Table 2b  
Regression of Achievement on Reading Frequency and SES

Variable	b	se	$\Delta R^2$	95% CI	F	t
SES	-12.66	5.16	.426	10.32, 42.26		-2.45*
Reading Freq.			.007		0.05	
B	-1.75	6.32		14.62, 52.52		-0.28
C	-0.00	6.32		-6.36, 33.16		0.00
Intercept	55.58	5.16		9.91, 42.33		10.77*

Note:  $R^2 = .43$ , adj.  $R^2 = .22$ ,  $F_{3,8} = 2.04$ ,  $MSE = 79.89$ ,  $n = 12$ .  $\Delta R^2$  represents the semi-partial correlation or the increment in  $R^2$  due to adding the respective variable. SES is a dummy variable (free/reduced lunch = 1, otherwise = 0), as are B (1, 0 = other) and C (1, 0 = other).

\* $p < .05$ .

Table 2c  
Comparisons of Achievement among Reading Frequency

Contrast	Estimated Mean Difference	Standard Error of Difference	.95CI
B vs. A	-1.75	6.32	-16.32, 12.82
C vs. A	-0.00	6.32	-14.57, 14.57
B vs. C	-1.75	6.32	-16.32, 12.82

\* $p < .05$ .

[Note – the above comparison represents the unadjusted comparisons (no Bonferroni corrections); these numbers obtained from regression output.]

Only SES was statistically related to achievement scores, with those receiving free for reduced lunch scoring about 12 to 13 points lower than those not receiving free/reduced lunch, on average. There were no statistical differences observed among the three levels of reading frequency.

Bonferroni and Scheffe adjusted confidence intervals are reported below.

```
. regr achievement i.read_freq_num i.ses_num
. pwcompare read_freq_num, bonf
```

Pairwise comparisons of marginal linear predictions

```
-----
```

	Contrast	Std. Err.	Bonferroni [95% Conf. Interval]	
read_freq_num				
2 vs 1	-1.75	6.320436	-20.81093	17.31093
3 vs 1	7.07e-15	6.320436	-19.06093	19.06093
3 vs 2	1.75	6.320436	-17.31093	20.81093

```
-----
```

```
. pwcompare read_freq_num, sch
```

```
-----
```

	Contrast	Std. Err.	Scheffe [95% Conf. Interval]	
read_freq_num				
2 vs 1	-1.75	6.320436	-20.62467	17.12467
3 vs 1	7.07e-15	6.320436	-18.87467	18.87467
3 vs 2	1.75	6.320436	-17.12467	20.62467

```
-----
```

(3) Results for GPA analysis.

*Table 3a*  
*Descriptive Statistics for GPA, SES, and Behavioral Problems*

Variable	Correlations			
	GPA	High	Mid	Behavior
GPA	---			
High SES	.78*	---		
Mid. SES	-.07	-.53	---	
Behavior	-.46	-.10	-.62*	---
Mean	2.56	0.39	0.31	0.46
SD	0.74	0.51	0.48	0.52

*Note:* High (1, 0 = otherwise) and Mid. SES (1, 0 = otherwise) are dummy variables, as is behavior (1 for problems, 0 = otherwise); n = 13.

*Table 3b*  
*Regression of GPA on Behavioral Problems and SES*

Variable	b	se	$\Delta R^2$	95%CI	F	t
Behavioral Prob.	-.27	.37	.01	-1.10, 0.57		-0.72
SES			.56		11.31*	
High	1.34*	.35		0.54, 2.13		3.81*
Mid	.46	.47		-0.60, 1.51		0.98
Intercept	2.03	.42		1.08, 2.98		4.83*

*Note:*  $R^2 = .78$ , adj.  $R^2 = .70$ ,  $F_{3,9} = 10.35^*$ ,  $MSE = 0.164$ , n = 13.  $\Delta R^2$  represents the semi-partial correlation or the increment in  $R^2$  due to adding the respective variable. High (1, 0 = otherwise) and Mid. SES (1, 0 = otherwise) are dummy variables, as is behavior (1 for problems, 0 = otherwise).

\*p < .05.

*Table 3c*  
*Comparisons of Achievement among Reading Frequency*

Contrast	Estimated Mean Difference	Standard Error of Difference	95% CI of Mean Difference
High vs. Low	1.34*	0.35	0.54, 2.12
Mid vs. Low	0.46	0.47	-.60, 1.51
High vs. Mid	0.88	0.31	.18, 1.58

\*p < .05.

Only SES was statistically related to GPA, with those in the high SES group showing statistically higher GPAs than either the middle or low SES groups. There was no statistical difference between the middle and low SES groups, nor was behavioral problem associated with GPA.

[Table 3c above are the unadjusted comparisons, Table 3d below contains the Bonferroni adjusted comparisons using the estimated means with behavioral problems mean used as 0.46 to obtained predicted means for each of the three SES groups.]

[Again note that the Excel spreadsheet se are too small and erroneous, so use tabled Bonferroni critical t and calculate CI using regression se.]

*Table 3c*  
*Comparisons of Achievement among Reading Frequency*

Contrast	Estimated Mean Difference	Standard Error of Difference	Bonferroni Adjusted 95% CI
High vs. Low	1.34*	0.35	0.31, 2.36
Mid vs. Low	0.46	0.47	-0.91, 1.83
High vs. Mid	0.88	0.31	-0.03, 1.79

\*p<.05, where p-values are adjusted using the Bonferroni method.

[Bonferroni critical t = 2.923 (3 comparions, 9 df)]

Scheffé confidence intervals are reported below.

*Table 3c*  
*Comparisons of Achievement among Reading Frequency*

Contrast	Estimated Mean Difference	Standard Error of Difference	Bonferroni Adjusted 95% CI
High vs. Low	1.34*	0.35	0.31, 2.36
Mid vs. Low	0.46	0.47	-0.91, 1.82
High vs. Mid	0.88	0.31	-0.02, 1.78

\*p<.05, where p-values are adjusted using the Scheffé method.