10

ANCOVA

Models with Continuous and Categorical Predictors

The material covered in this chapter represents an integration of the material covered in Chapters 6 and 9. In Chapter 6 we considered models with continuous predictors; Chapter 9 was concerned with models with categorical predictors. We will now consider models that include both sorts of predictors. Such models are more traditionally covered under the heading of analysis of covariance (ANCOVA). We feel that this is a confusing label for these models and prefer to think of them simply as models involving both categorical and continuous predictor variables.

In Chapter 7, we saw that an extension of models with continuous predictors included products of those predictors, or interactions among them, as predictors as well. Similarly, in the last chapter we considered as predictors interactions among categorical predictor variables. By extension, we will also consider in this chapter models that include not only categorical and continuous predictors but also interactions between the two kinds of predictor variables.

Interest in these sorts of models originally developed within the tradition of methods for the analysis of experimental and quasi-experimental research designs. In this tradition, researchers were primarily interested in the effects of experimental factors, or the categorical predictor variables, and wished to estimate those effects while controlling for some continuously measured concomitant variable, called a *covariate*. As is detailed below, this control might be desired because of the increase in power that such control might produce or because the concomitant variable was redundant with the categorical ones. Models involving both categorical and continuous predictor variables have, however, a wide range of application outside of experimental research designs. For instance, in sociology or political science, one might be interested in the effects of both a categorical variable (e.g., sex) and a continuous one (e.g., personal income) on some dependent variable, and might wish to estimate each of these effects when controlling for the other. In other words, unlike in the experimental design domain, the focus in these models need not be on the effects of the categorical variables and what happens to those effects when a continuous predictor variable is controlled. We might be just as interested in the effects of the continuous predictor variable and how those effects change when we control for a categorical one.

Given the historical tradition of these models within the context of experimental design, we start the chapter by illustrating the use of models with both categorical and continuous predictors within the context of an experimental design, where the primary interest is in the effects of the categorical variables. We use this context in order to illustrate the reasons we might want to control for a continuously measured concomitant variable. In later sections of the chapter, we illustrate the use of these models in a context where the researcher's interest is primarily in the effects of the continuous predictor when a categorical one is controlled.

CONTROLLING FOR AN ORTHOGONAL CONTINUOUS VARIABLE IN A FACTORIAL DESIGN

Suppose we were evaluating a curriculum innovation in a secondary school. Students were randomly assigned to either the new or the old curriculum. In addition, each curriculum was taught by two different teachers, and students were assigned to one of the two teachers on a random basis. Thus, the experimental design is a two-factor crossed design, with two levels of both the curriculum and teacher factors. Ten students have been randomly assigned to each of the resulting four conditions. The dependent variable is the student's score on a standardized achievement test given at the end of the curriculum. In addition, we have a pretest measure from each student, indicating his or her achievement in the domain in question prior to exposure to either the new or old curriculum. The hypothetical raw data are given in Figure 10.1. The variable Z_i represents each student's pretest achievement score. The X_{1i} variable is a contrast-coded variable that codes curriculum (–1 if Old Curriculum; 1 if New Curriculum); X_{2i} is a contrast-coded variable that codes teacher (–1 if Teacher A; 1 if Teacher B); and Y_i represents each student's post-test achievement score.

The means for both Z_i and Y_i for each of the four cells of the design are given in Figure 10.2. Notice that the data have been constructed so that all four pretest means equal 50. Since students have been randomly assigned to the four treatment conditions

FIGURE 10.1 Hypothetical experimental data

Y_i	X_{1i}	X_{2i}	Z_i	Y_i	X_{1i}	X_{2i}	Z_i
58	1	–1	50	57	1	–1	49
63	1	–1	49	61	1	–1	52
65	1	–1	53	57	1	–1	50
56	1	–1	47	67	1	–1	51
60	1	–1	53	56	1	–1	46
50	–1	–1	49	62	–1	–1	54
58	–1	–1	51	55	–1	–1	48
52	–1	–1	50	63	–1	–1	52
55	–1	–1	47	50	–1	–1	46
57	–1	–1	52	58	–1	–1	51
61	1	1	47	59	1	1	49
71	1	1	53	65	1	1	54
68	1	1	52	60	1	1	46
58	1	1	48	65	1	1	51
68	1	1	51	65	1	1	49
47	–1	1	46	62	–1	1	51
56	–1	1	51	51	–1	1	49
63	–1	1	53	54	–1	1	51
53	–1	1	48	58	–1	1	50
52	–1	1	47	54	–1	1	54

after they took the pretest, we would expect their mean pretest scores to be very similar, although the chances are that they would not all be identical. (They have been constructed to be identical here for didactic purposes.)

The reason for constructing these data with equal pretest means is that as a result the pretest is not correlated with condition. To see this, suppose we regressed the pretest Z_i on X_{1i}, X_{2i}, and their interaction. With all three contrast codes as predictors, the predicted values of Z_i would equal the four cell means. Each of these cell means also equals the grand mean. Hence, a model with all three contrast-coded predictors generates the same predicted values for every case as a simple model in which we predict the grand mean for every case. As a result, condition is entirely unrelated to pretest.

FIGURE 10.2 Pretest and post-test means by condition

Condition	$\bar{Z}_k$	$\bar{Y}_k$
Old Curriculum, Teacher A	50	56
Old Curriculum, Teacher B	50	55
New Curriculum, Teacher A	50	60
New Curriculum, Teacher B	50	64

ANOVA of the Data

Let us conduct a straightforward two-way ANOVA on the post-test data, ignoring the pretest variable for the moment. We regress Y_i on X_{1i}, X_{2i}, and X_{3i}, where X_{3i} is the interaction contrast code, formed by multiplying X_{1i} and X_{2i} together. The resulting model is:

$$\hat{Y}_i = 58.75 + 3.25X_{1i} + 0.75X_{2i} + 1.25X_{3i}$$

with a sum of squared errors of 710. The value of the regression coefficient for X_{1i}, 3.25, equals half the difference between the average of the two mean values of Y_i under the new curriculum and the average of the two mean values of Y_i under the old curriculum. The regression coefficient for X_{2i}, 0.75, equals half the difference between the average under Teacher B and the average under Teacher A. The regression coefficient for the interaction contrast code, 1.25, equals half the difference between the average of the Old Curriculum/Teacher A and New Curriculum/Teacher B conditions and the average of the Old Curriculum/Teacher B and New Curriculum/Teacher A conditions. Conceptually it tells us that the curriculum difference is larger under Teacher B than under Teacher A.

We can calculate the sum of squares due to each contrast-coded predictor by computing a series of compact models that omit each contrast code in turn. Alternatively, we can use the formula for the SSR associated with a contrast-coded predictor that we have used before:

$$\text{SSR} = \frac{\left(\sum_k \lambda_k \bar{Y}_k\right)^2}{\sum_k \frac{\lambda_k^2}{n_k}}$$

Accordingly, the sum of squares explained by X_{1i} equals 422.5, that explained by X_{2i} equals 22.5, and that explained by X_{3i} equals 62.5. Since we have 10 observations in each of the four conditions, the three predictors are nonredundant and hence these sums of squares add up to the between condition sum of squares. The ANOVA source table for this analysis is presented in Figure 10.3.

FIGURE 10.3 Two-way ANOVA source table

Source	*b*	*SS*	*df*	*MS*	*F*	*PRE*
Between or model		507.5	3	169.17	8.58	.42
Curriculum	3.25	422.50	1	422.50	21.42	.37
Teacher	0.75	22.50	1	22.50	1.14	.03
Curriculum x Teacher	1.25	62.50	1	62.50	3.17	.08
Error		710.00	36	19.72		
Total		1217.50	39			

Including an Orthogonal Covariate

In this analysis, we have made no use of the pretest Z_i. We might decide to add it as a predictor to the model for a number of reasons. For instance, we might be interested in asking how curriculum and teacher affect post-test performance when we control for the pretest or when we look within levels of the pretest. In other words, we might be interested in these effects over and above differences in performance that existed at the time of the pretest. A seemingly different reason for including it as a predictor in the model is that we might expect it to be highly correlated with the post-test, since presumably it is only an earlier version of the post-test, measuring the same domain of achievement. If this is so, then, as we shall see, it might make our tests of curriculum and teacher effects, and their interaction, considerably more powerful.

In these data, the pretest scores are highly related to the post-test scores. The sum of squared errors from a simple regression model in which Y_i is regressed on Z_i equals 855.57. From the last or "Total" row of the source table of Figure 10.3 we see that the sum of squares of Y_i for the simplest single-parameter model equals 1217.5. Accordingly, a comparison between the two-parameter simple regression model using Z_i to predict Y_i and a one-parameter model yields a PRE of .297 and $F_{1,38} = 16.05$. Thus, the relationship between the pretest and the post-test is substantial and highly reliable.

As we have seen, the pretest Z_i is uncorrelated with condition, since all of the pretest conditions means are identical. Accordingly, the sum of squares of Y_i that could be explained by Z_i will not overlap with the sums of squares of Y_i explained by the three condition contrast codes. Because the sum of squares explained by these three contrast-coded predictors equals 507.5, the sum of squares explained by Z_i equals 361.93, and Z_i is nonredundant with the three contrast-coded predictors, the sum of squares explained by the pretest, Z_i, *plus* the three contrast-coded predictors ought to equal 507.5 + 361.93 or 869.43. Accordingly, the sum of squared errors for a model in which Y_i is regressed on Z_i X_{1i}, X_{2i}, and X_{3i} should equal 1217.5 – 869.43 or 348.07.

In the ANOVA source table of Figure 10.3, the denominator for each of the F statistics testing the significance of the condition differences equals the mean square error within or the mean square error from the final augmented model that includes all predictor variables. If we include the pretest Z_i as an additional predictor, this mean square error ought to be reduced substantially due to the fact that the sum of squares potentially explainable by Z_i has been controlled for or removed from the sum of squared errors. Thus, by including Z_i as an additional predictor in the model, we would expect our tests of condition differences in Y_i to be more powerful, yielding larger F values.

The resulting augmented model, with both the pretest and the three contrast-coded predictors, is estimated as:

$$\hat{Y}_i = -4.52 + 1.27Z_i + 3.25X_{1i} + 0.75X_{2i} + 1.25X_{3i}$$

with a sum of squared errors of 348.07.

Notice that the coefficients in this model for the three contrast-coded predictors have not changed from the model where Z_i was not included as a predictor. Since Z_i is nonredundant with all of the contrast-coded predictors, its inclusion has no effect on the value of their coefficients or on their interpretation. The coefficient for X_{1i} continues to tell us about the magnitude of the difference in Y_i due to curriculum. The coefficient for X_{2i} continues to tell us about the magnitude of the mean teacher difference. And the coefficient for X_{3i} continues to inform us about the degree to which the curriculum difference is larger under Teacher B than under Teacher A.

We could test whether the coefficient for Z_i significantly differs from zero by comparing this augmented model with the model presented earlier that included only the contrast-coded predictors. The value of PRE that results from this comparison is:

$$\text{PRE} = \frac{710.00 - 348.07}{710.00} = .510$$

with an associated F statistic of

$$F_{1,35} = \frac{.510/1}{(1 - .510)/35} = 36.39$$

Accordingly, we can conclude that, independent of condition, or on average within condition, pretest scores significantly relate to post-test scores.

We can test whether the condition means differ by testing the regression coefficients in this model for each of the three contrast-coded predictors. Let us start with the omnibus test of whether there are any differences in the condition means of Y_i. To do this test, we want to compare the augmented model that includes Z_i and the three contrast-coded predictors with a compact one that includes only Z_i. We have already said that the sum of squared errors from this compact simple-regression model equals 855.57. Accordingly, the value of PRE for the omnibus test is:

$$\text{PRE} = \frac{855.57 - 348.07}{855.57} = .593$$

And the omnibus F statistic, having 3 and 35 degrees of freedom, is:

$$F_{3,35} = \frac{.593/3}{(1 - .593)/35} = 17.01$$

An equivalent expression for the F statistic is given in terms of the sums of squares:

$$F_{3,35} = \frac{507.50/3}{348.07/35} = 17.01$$

where 507.50 is the reduction in the sum of squared errors as we move from the compact to the augmented model, and 348.07 is the sum of squared errors of the augmented model.

Notice that this F statistic is nearly twice as large as the F statistic for the omnibus test based on the augmented model that did not include the pretest as a predictor (given in Figure 10.3). The reason for this difference is that the sum of squared errors from the augmented model is now considerably less than it was without the pretest included. It is also true that the degrees of freedom for error have been reduced by 1 as a result of the additional parameter for the pretest estimated in the augmented model. In combination, however, the substantially smaller sum of squared errors and the slightly smaller degrees of freedom for error result in a considerably smaller mean square error.

In other words, the denominator of the F ratio for the omnibus test of condition differences when the pretest is included equals 348.07/35 or 9.94. In the model that did not include the pretest, the denominator of the F ratio for the omnibus test of condition differences equaled 710.0/36 or 19.72. The net result of including the pretest as a predictor, then, has been to reduce the mean square error and increase the power of tests of the condition effects. Such an increase in power will happen whenever there is a significant relationship between the continuously measured predictor variable, in this case the pretest, and the dependent variable in the augmented model that includes the categorical predictor variables. The full source table for the model that includes the pretest and all three contrast-coded predictors is given in Figure 10.4. Let us compare this ANCOVA source table with the ANOVA source table of Figure 10.3.

First, notice that the test of the overall model—comparing this model that includes four predictor variables with the single parameter model that includes no predictors—yields substantially higher values of both PRE and F than the test of the overall model that did not include the pretest. These larger values are entirely attributable to the fact that the pretest is highly related to the dependent variable. Accordingly, the sum of squares attributable to the model, 869.43, has dramatically increased from the ANOVA table, while the error sum of squares has dramatically decreased. The decrease in the sum of squared errors for the model, and equivalently the increase in the sum of squares explained by the model, is exactly equal to the sum of squares associated with the pretest, i.e., 361.93.

Second, notice that the rows for the three contrast-coded variables, representing curriculum, teacher, and their interaction, have the same sums of squares, degrees of freedom, and mean squares as they did in the ANOVA source table. Because the pretest is unrelated to curriculum, teacher, and their interaction, neither the regression coefficients for these contrast-coded predictor variables nor their sums of squares are affected by the inclusion of the pretest in the model. The inclusion of the pretest, however, does have a major effect on the row in the source table referring to error in the model, as just described. As a result, all of the F statistics used to test the teacher effect, curriculum effect, and the interaction effect on the post-test dependent variable are substantially larger than they were in the ANOVA source table of Figure 10.3. Whereas the curriculum $\times$ teacher interaction was not significant in the ANOVA source table, it now is. The positive coefficient associated with this interaction contrast code tells us that the new–old curriculum difference, while significant on average across Teachers, is significantly larger for Teacher B than it is for Teacher A.

This analysis has illustrated one of the major reasons for including a continuously measured predictor variable (equivalently called a covariate) in randomized experimental research designs. If that predictor variable is measured prior to randomization of participants to conditions, on average it will be unrelated to the contrast-coded variables

FIGURE 10.4 ANCOVA source table

Source	*b*	*SS*	*df*	*MS*	*F*	*PRE*
Model		869.43	4	217.36	21.87	.71
Pretest	1.27	361.93	1	361.93	36.39	.51
Between conditions		507.50	3	169.17	17.01	.59
Curriculum	3.25	422.50	1	422.50	42.48	.55
Teacher	0.75	22.50	1	22.50	2.26	.06
Curriculum x Teacher	1.25	62.50	1	62.50	6.28	.15
Error		348.07	35	9.94		
Total		1217.50	39			

that code experimental condition. The purpose of including such a variable is to increase the power of the analysis that tests for condition differences in the dependent variable. If the covariate is in fact unrelated to condition, then neither the regression coefficients for the various contrast-coded predictors that code condition differences nor their sums of squares will be affected by its inclusion in the model. The null hypothesis associated with the test of a given contrast-coded predictor will also not change. Regardless of the inclusion of the covariate, we will still be testing for differences among the condition means on the dependent variable, as coded by the λ values. Tests for condition differences will be more powerful as a result of including a covariate whenever the test of whether the covariate's regression coefficient in the full augmented model differs from zero yields a significant F statistic. When the covariate is unrelated to the dependent variable, then the decrease in the sum of squared errors resulting from the inclusion of the covariate will not offset the decrease of the degrees of freedom for error in the model. The ideal covariate, therefore, in this situation, is one that is as highly associated as possible with the dependent variable controlling for the categorical variables or within levels of the categorical variables.

Even with random assignment of participants to condition after measuring the covariate, it will almost never be the case that the covariate will be entirely nonredundant with the condition contrast codes. In other words, it will be a very rare event for all of the pretest or covariate means in the various experimental conditions to be exactly equal. Our example, then, is obviously a constructed one, designed simply to illustrate what happens in the pure case, when the covariate is completely independent of condition. In any given study, there will in all probability be some nonsignificant relationships between the covariate and the contrast codes that represent condition. Nevertheless, the inclusion of a covariate will increase the statistical power of tests of condition differences, given a covariate that is significantly related to the dependent variable within levels of the categorical variables.

As we said in the introduction to this chapter, within the context of experimental designs the usual interest in including a continuously measured predictor variable with a set of categorical ones is to examine what happens to tests of condition differences when we control for the continuously measured covariate. As we have seen, with a covariate measured prior to random assignment of participants to condition, the result will generally be an increase in statistical power for tests of condition differences. There is no necessary reason, however, for confining our interpretations of the model that includes both kinds of predictors to this typical interest. In other words, there is nothing to prevent us from turning the interpretation of this model around—concentrating not on the tests of mean differences while controlling for the covariate, but on a test of the pretest–post-test relationship while controlling for condition differences on the post-test. If we simply regress the post-test on the pretest, the pretest's regression coefficient equals 1.27. The sum of squared errors for this simple regression model equals 855.57. A test of the simple pretest–post-test relationship yields a PRE of .297 and an F of 16.05 with 1 and 38 degrees of freedom. When we examine the pretest–post-test relationship controlling for the categorical variables, as given in the ANCOVA source table of Figure 10.4, the pretest's coefficient is still 1.27, but a test of whether it is reliably related to the post-test within condition levels yields a PRE of .51 and an F of 36.39 with 1 and 35 degrees of freedom. Thus, we might equivalently look at this analysis as a way of increasing the power of tests of the pretest–post-test relationship by controlling for experimental conditions.

ANALYSIS OF POST-TEST–PRETEST DIFFERENCE SCORES

In giving a rationale for the analysis that includes pretest as a predictor, we suggested that it might make sense to examine the effects of curriculum and teacher on the post-test when controlling for pretest differences or holding constant pretest performance. It might seem that an equivalent way of doing this analysis would be to examine condition differences in improvement from the pretest to the post-test. To do such an analysis, we might logically compute a new dependent variable equal to $Y_i - Z_i$, assuming they were both measures of the same thing—the pretest taken before the experiment and the post-test at its conclusion. This *difference score* tells us about each individual's improvement in achievement during the course of the study. We would then be interested in condition differences in the mean $Y_i - Z_i$ difference scores. Since all Z_i condition means are identical, the mean differences among conditions on the $Y_i - Z_i$ difference scores will be equivalent to the mean differences among conditions on Y_i.

To examine condition effects on this improvement difference score, let us regress it on the three contrast-coded predictors that define condition. The following estimated parameters result:

$$\widehat{Y_i - Z_i} = 8.75 + 3.25X_{1i} + 0.75X_{2i} + 1.25X_{3i}$$

with a sum of squared errors of 364.00. The intercept in this model equals the average of the condition means on the difference score or, equivalently, the difference between the means of the condition means of Y_i and of the condition means of Z_i. Somewhat surprisingly, perhaps, the regression coefficients for the three contrast-coded predictors have not changed as a result of changing the dependent variable to the $Y_i - Z_i$ difference score. Both in this difference score analysis and in the analysis where the post-test was the dependent variable and the pretest was included as a predictor, the regression coefficients continue to equal what they did in the simple analysis of variance with Y_i as the dependent variable and no pretest. This invariance is once again due to the fact that the pretest is uncorrelated with condition.

To illustrate algebraically why these regression coefficients have not changed, let us examine the algebraic expression for the regression coefficients associated with contrast-coded predictors in this difference score analysis. Using our standard formula, but substituting the difference score means for the usual $\bar{Y}$ values, we get:

$$\frac{\sum_k \lambda_k (\overline{Y_k - Z_k})}{\sum_k \lambda_k^2} = \frac{\sum_k \lambda_k (\bar{Y}_k - \bar{Z}_k)}{\sum_k \lambda_k^2} = \frac{\sum_k \lambda_k \bar{Y}_k - \sum_k \lambda_k \bar{Z}_k}{\sum_k \lambda_k^2}$$

Since all $\bar{Z}_k$ are identical, the expression $\sum \lambda_k \bar{Z}_k$ equals zero and this expression for the regression coefficient for each of the predictors reduces to what it is when simply Y_i is the dependent variable in the model. In sum, these regression coefficients equal the coded condition differences in the mean difference scores ($\bar{Y}_k - \bar{Z}_k$), which, given equal $\bar{Z}_k$, are equivalent to the coded differences in $\bar{Y}_k$.

This difference score analysis gives us the difference score ANOVA source table of Figure 10.5. Notice that just as the regression coefficients for the contrast-coded predictors in this difference score analysis equal what they were in the analysis that included pretest as a predictor variable, so too are their sums of squares equal to what they have been all along. Once again, this equivalence is due to the fact that Z_i is uncorrelated with condition.

This analysis, however, is different from both the ANOVA and the ANCOVA in terms of the sum of squares for error and the sum of squares total. The total sum of squares is now equal to the sum of squared variation in the $Y_i - Z_i$ difference scores, which in this case is less than the total sum of squares in Y_i. Since this total sum of squares has been reduced and since the sums of squares explained by the three contrast-coded predictors are unchanged, the error sum of squares in this analysis must be less than the sum of squared errors in the earlier analysis of variance with Y_i as the dependent variable. As a result, the F and PRE statistics for the omnibus test of any condition differences and for the three single-degree-of-freedom tests are all larger than they were in the original ANOVA source table of Figure 10.3.

Notice, however, that the sum of squared errors in this difference score source table is larger than the sum of squared errors was in the analysis that included the pretest as a predictor variable (ANCOVA source table of Figure 10.4). As a result, the F and PRE statistics testing condition differences in this difference score analysis are all slightly smaller than they were in the source table of Figure 10.4. In sum, while this difference score analysis is more powerful in this case than the simple ANOVA of Y_i, it is not as powerful as the ANCOVA in which the dependent variable was Y_i and Z_i was included as a predictor variable.

FIGURE 10.5 Difference score ANOVA source table

Source	*b*	*SS*	*df*	*MS*	*F*	*PRE*
Between or model		507.50	3	169.17	16.73	.58
Curriculum	3.25	422.50	1	422.50	41.79	.54
Teacher	0.75	22.50	1	22.50	2.22	.06
Curriculum x Teacher	1.25	62.50	1	62.50	6.18	.15
Error		364.00	36	10.11		
Total		871.50	39			

To understand why this is so, let us examine the difference score model:

$$Y_i - Z_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

We can re-express this model by adding Z_i to both sides of the equation:

$$Y_i = \beta_0 + Z_i + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

This difference score model now looks very similar to the ANCOVA model in which the pretest was used as a predictor variable. There is, however, one major difference. Instead of estimating a parameter for the pretest variable as we did in the ANCOVA model, we have set the parameter value equal to 1.0. By doing the difference score analysis, we have in effect assumed that the parameter value for the pretest equals 1.0, rather than letting it be a free parameter and deriving its least-squares estimate.

By definition, the least-squares estimates are those that minimize the sum of squared errors. Accordingly, the sum of squared errors in a model where the coefficient for the covariate is fixed at 1.0 cannot be less than the sum of squared errors in the ANCOVA model, since in the latter model the coefficient for the covariate is the least-squares estimate. Hence, the difference score analysis will generally be less powerful than the ANCOVA. Frequently, it will be substantially less powerful and may, in fact, be even less powerful than the simple ANOVA model.

Once we realize that this difference score model is identical to the ANCOVA model, except that we have fixed the coefficient for the covariate at 1.0 instead of estimating it, we can rewrite the difference score analysis source table as we have in Figure 10.6. In this revised table, the sum of squares total refers to the total sum of squares of Y_i rather than the total sum of squares in the $Y_i - Z_i$ difference score.

FIGURE 10.6 Revised difference score ANOVA source table

Source	*b*	*SS*	*df*	*MS*	*F*	*PRE*
Model		853.50	3	284.5	28.14	.70
Pretest	1.00	346.0	0			
Between conditions		507.50	3	169.17	16.73	.58
Curriculum	3.25	422.50	1	422.50	41.79	.54
Teacher	0.75	22.50	1	22.50	2.22	.06
Curriculum x Teacher	1.25	62.50	1	62.50	6.18	.15
Error		364.00	36	10.11		
Total		1217.50	39			

Notice in this source table that the degrees of freedom for the pretest, Z_i, equal zero since in this model its coefficient has been set at 1.0 rather than estimated from the data. The sum of squares associated with Z_i equals the difference between the sum of squares total for Y_i and the sum of squares total for the $Y_i - Z_i$ difference score.

In these data, the difference score analysis is only slightly less powerful than the ANCOVA. This near equivalence results from the fact that the estimated parameter for the covariate in the ANCOVA model, 1.27, is rather close to the value of 1.0 at which it is fixed in the difference score analysis. We could test whether the ANCOVA model results in significantly smaller errors of prediction than the difference score model. This is equivalent to testing whether the parameter associated with the covariate in the augmented ANCOVA model significantly differs from 1.0. For this test, the ANCOVA model, in which the covariate's parameter is estimated, is the augmented one:

$$\text{MODEL A: } Y_i = \beta_0 + \beta_1 Z + \beta_2 X_{1i} + \beta_3 X_{2i} + \beta_4 X_{3i} + \varepsilon_i$$

The compact model is the difference score model in which the pretest's parameter is fixed at 1.0:

$$\text{MODEL C: } Y_i = \beta_0 + Z + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$

Comparing the sums of squared errors for these two models, we get:

$$\text{PRE} = \frac{364.00 - 348.07}{364.00} = .044$$

which converts to an F of 1.61 with 1 and 35 degrees of freedom. Hence, in these data, we cannot conclude that the parameter for the covariate is significantly different from 1.00. This conclusion means that the ANCOVA model for these data is not significantly more powerful than the difference score analysis.

Having said that the ANCOVA model will generally be more powerful than the difference score analysis, we should note that the ANCOVA uses up a degree of freedom in estimating the covariate's parameter while the difference score analysis does not. It is thus possible, if the estimated slope for the covariate is close to 1.00, for the mean square error from the difference score model to be smaller than that from the ANCOVA model. The difference score analysis may also be easier than the ANCOVA to describe to others who are untrained statistically.

THE CASE OF A PARTIALLY REDUNDANT CONTINUOUS PREDICTOR

So far we have been illustrating one major reason for including a continuously measured concomitant variable in an analysis of condition differences. By including a covariate that is measured prior to randomization of participants to conditions, the covariate will generally be unrelated to condition and its inclusion in the model will increase the power of the tests of mean condition differences, so long as the covariate is highly related to the dependent variable of interest.

But there are other important reasons for including a continuously measured covariate in an analysis of mean condition differences. These reasons arise precisely *because* the covariate is correlated with condition and we wish to examine the condition differences over and above, or controlling for, covariate differences. There are two primary occasions when this is of interest. Analytically, these occasions are identical—we simply will be estimating an ANCOVA model with a covariate that is partially redundant with the categorical predictor(s)—but they differ in important theoretical ways that give rise to rather different interpretations.

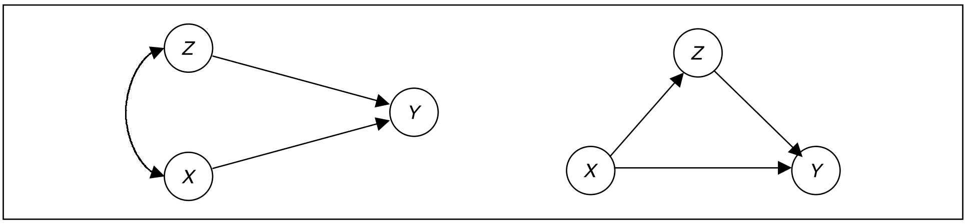
The first occasion on which controlling for a partially redundant covariate is of interest occurs when, for whatever reason, there are pre-existing differences among participants in the various conditions of the design, and the researcher wishes to attempt to examine the effects of the categorical variable(s) free from those differences. Suppose, for instance, in the example we have been using, that students had not been randomly assigned to the four cells defined by the teacher and curriculum factors. If the decision about which students got which teacher–curriculum condition was not based on a random decision rule, but was instead based on some unknown assignment rule (e.g., keeping last year's classes intact), then we might expect differences among the various conditions on the pretest long before the students had been exposed to the teacher or to the curriculum. We would then quite reasonably want to control statistically for these pretest differences in looking at condition differences on the post-test. In other words, we might then like to control for the pretest in our analysis so that we could look at differences among conditions on the post-test free from the pre-existing differences on the pretest.

This use of ANCOVA to control for or adjust for known pre-existing differences among participants is particularly common in what have come to be known as quasi-experimental research designs where assignment to condition has not been on a purely random basis (Judd & Kenny, 1981a; Shadish, Cook, & Campbell, 2002). Although this adjustment function of ANCOVA (attempting to equate participants on pre-existing covariate differences) is widely used in such quasi-experimental designs, it is important to recognize that this analytic approach is not a general solution to the problem of causal inference in nonexperimental research. The ability to reach causal conclusions about the effects of various independent variables depends not on the statistical analysis that one undertakes but on the research design.

The second occasion on which it becomes important to conduct an ANCOVA with a covariate that is partially redundant with condition is when one is attempting to test a mediational model (Baron & Kenny, 1986; Judd & Kenny, 1981b). Suppose one has conducted an experiment in which the effect of some treatment difference on a dependent variable is to be estimated. If there is an effect of the experimental conditions (as estimated through an analysis of variance), then there certainly must be an underlying mediating mechanism or process that is responsible for that effect. Often one has an idea about the nature of that mechanism and one may attempt to measure one or more variables that intervene in a causal chain between the experimental manipulation and the ultimate dependent variable that it affects. These intervening variables are known as mediating variables. If, in fact, the hypothesized mediating mechanism is partially responsible for the effect of the experimental treatment on the ultimate dependent variable, then that effect should no longer be as apparent once one conducts an ANCOVA, controlling for the mediating variable.

Although the analytic models in these two cases are essentially identical—to determine if there are condition differences once one controls for another variable that is known to be correlated with condition—they represent different theoretical models about the underlying causal process linking the categorical independent variable(s) with the dependent variable. This difference is perhaps most clearly shown by representing the two cases in terms of hypothesized causal models, which are shown in Figure 10.7. In these causal models, single-headed arrows are used to represent hypothesized causal effects of one variable upon another, while curved double-headed arrows are simply meant to represent a correlation between two variables with no causal claims about the process responsible for that correlation. As previously, we use X to represent a categorical independent variable (here with only two levels, thus needing only one contrast-coded predictor), Z represents the continuously measured covariate, and Y represents the dependent variable.

FIGURE 10.7 Theoretical causal models for two sorts of partially redundant covariates



The first causal model is meant to represent the quasi-experimental case, in which participants have not been randomly assigned to the two levels of the categorical independent variable and therefore there are differences between the two groups on some continuously measured variable Z (X and Z are thus partially redundant with each other). Here one wishes to estimate the effect of X on Y free from the potential confounding influence of Z on Y. The goal then is to somehow come up with a more accurate estimate of the X effect, controlling for or eliminating any confounding due to the partially redundant covariate.

The second case is one in which random assignment to levels of the categorical independent variable has been accomplished, and one measures both the ultimate dependent variable, Y, as well as another variable, Z, that is thought to be affected by X. As the model indicates, here there are two ways in which X may exert an effect on Y: the first is through the mediating variable, Z, and the second is over and above that mediating variable. In this case, one controls for the mediating variable, Z, in an ANCOVA because one is interested is examining whether there is any residual direct effect of X on Y over and above the hypothesized mediating process through Z.

As we have said, the analytic approach in these two cases is identical—estimating effects while controlling for a continuously measured covariate that is partially redundant with the categorical independent variable. Nevertheless, to illustrate both and the differences in interpretations that result from the different causal models underlying them, we provide an example of each one.

Examining Partial Effects in a Quasi-Experimental Situation

In Figure 10.8, we present data that have been somewhat modified from those used in the example earlier in this chapter. The values of Y_i are identical to what we had earlier in Figure 10.1, as are the definitions of the contrast-coded predictor variables, with X_{1i} coding curriculum and X_{2i} coding teacher. What we have modified slightly are the values of the pretest variable, Z_i, so that now there are differences among the four cells of the design (resulting from the crossing of the two factors) on the pretest measure. The four cell means for both the pretest, Z_i, and the post-test, Y_i, are given in Figure 10.9.

There is now a definite relationship in these data between condition and Z_i, the pretest, since its condition means are no longer equal to each other. If we regressed Z_i on the three contrast-coded predictors (X_{3i} being defined as earlier to capture the curriculum – teacher interaction), we get the following parameter estimates:

FIGURE 10.8 Modified pretest–post-test data

Y_i	X_{1i}	X_{2i}	Z_i	Y_i	X_{1i}	X_{2i}	Z_i
58	1	−1	51	57	1	−1	50
63	1	−1	50	61	1	−1	53
65	1	−1	54	57	1	−1	51
56	1	−1	48	67	1	−1	52
60	1	−1	54	56	1	−1	47
50	−1	−1	48	62	−1	−1	53
58	−1	−1	50	55	−1	−1	47
52	−1	−1	49	63	−1	−1	51
55	−1	−1	46	50	−1	−1	45
57	−1	−1	51	58	−1	−1	50
61	1	1	50	59	1	1	52
71	1	1	56	65	1	1	57
68	1	1	55	60	1	1	49
58	1	1	51	65	1	1	54
68	1	1	54	65	1	1	52
47	−1	1	44	62	−1	1	49
56	−1	1	49	51	−1	1	47
63	−1	1	51	54	−1	1	49
53	−1	1	46	58	−1	1	48
52	−1	1	45	54	−1	1	52

FIGURE 10.9 Modified pretest and post-test means by condition

Condition	$\bar{Z}_k$	$\bar{Y}_k$
Old Curriculum, Teacher A	49	56
Old Curriculum, Teacher B	48	55
New Curriculum, Teacher A	51	60
New Curriculum, Teacher B	53	64

$$\hat{Z}_i = 50.25 + 1.75X_{1i} + 0.25X_{2i} + .75X_{3i}$$

with a sum of squared errors of 225.94. The estimated slopes in this model tell us about the differences in the mean pretest values, according to the following frequently used expression for the slopes of contrast-coded predictor variables:

$$\frac{\sum_k \lambda_k \bar{Z}_k}{\sum_k \lambda_k^2}$$

To test whether these pretest means are significantly different from each other we could compare this as Model A with a Model C that simply predicted the grand mean of Z_i for every observation. This Model C has a sum of squared errors of 373.5, resulting in a PRE of .40 and an F of 7.84 with 3 and 36 degrees of freedom. Thus, there is significant redundancy between the pretest and the categorical variables that represent the conditions in which an observation is observed.

Turning to the analysis of post-test scores, if we conducted that analysis whilst ignoring the pretest, the exact same model would result as before:

$$\hat{Y}_i = 58.75 + 3.25X_{1i} + 0.75X_{2i} + 1.25X_{3i}$$

and the ANOVA source table that was given earlier (Figure 10.3) would continue to be found, since we have done nothing to alter the Y_i variable in this modified dataset.

However, the ANCOVA model, controlling for the partially redundant pretest variable, looks very different from what it was previously:

$$\hat{Y}_i = -4.84 + 1.26Z_i + 1.04X_{1i} + 0.43X_{2i} + 0.30X_{3i}$$

$$\text{SSE} = 348.07$$

Note that the regression coefficients for the contrast-coded predictors have been dramatically affected by the inclusion of the pretest variable Z_i. No longer does the coefficient for X_{1i} equal half the difference between the average Y_i under the new and old curricula. Similarly, the coefficients for X_{2i} and X_{3i} can no longer be interpreted as they were previously in terms of differences among various $\bar{Y}_k$. In other words, with the inclusion of a covariate that is partially redundant with the contrast-coded predictors, the regression coefficients for the contrast-coded predictors are no longer equal to:

$$\frac{\sum_k \lambda_k \bar{Y}_k}{\sum_k \lambda_k^2}$$

and therefore a test of whether the parameters for these contrast-coded predictors equal zero is no longer a simple test of a comparison among the Y_i means in the various conditions.

With the inclusion of a covariate, the regression coefficient for a contrast-coded predictor variable is equal to:

$$\frac{\sum_k \lambda_k \bar{Y}_k}{\sum_k \lambda_k^2} - b_z \frac{\sum_k \lambda_k \bar{Z}_k}{\sum_k \lambda_k^2}$$

where b_z is the regression coefficient associated with the covariate in the full model that includes both the covariate and the set of contrast-coded predictors.

To illustrate this expression, let us calculate the value of the regression coefficient for X_{1i} in the model that includes Z_i. Notice that the first half of this expression equals the regression coefficient for X_{1i} in the model that did not include the covariate, i.e., 3.25. The second half of the expression equals the parallel difference coded by the contrast weights among the *covariate or pretest* condition means, weighted by the regression coefficient for the pretest. Numerically, for the coefficient for X_{1i} the second half of this expression equals:

$$b_z \frac{\sum_k \lambda_k \bar{Z}_k}{\sum_k \lambda_k^2} = 1.26\left(\frac{53 + 51 - 48 - 49}{4}\right) = 1.26(1.75) = 2.21$$

In sum, then, according to this expression the regression coefficient for the X_{1i} contrast-coded predictor equals:

$$3.25 - 1.26(1.75) = 1.04$$

This new expression for the regression coefficient for a contrast-coded predictor in the presence of a covariate is readily interpreted. It is equal to the magnitude of the difference among the $\bar{Y}_k$ coded by the contrast weights, adjusting for or subtracting off the magnitude of the same difference among the covariate condition means $\bar{Z}_k$. The degree to which the coded comparison among the $\bar{Y}_k$ is adjusted by the same comparison among the $\bar{Z}_k$ depends on the magnitude of the covariate's regression coefficient, b_z. In sum, the regression coefficients for contrast-coded predictors in the presence of a covariate tell us about the differences among the $\bar{Y}_k$ coded by the contrast weights, adjusting that difference for the parallel difference that exists in the covariate condition means $\bar{Z}_k$. The degree to which this adjustment is performed depends on the magnitude of the within-condition relationship between the covariate and the dependent variable, i.e., the partial regression coefficient for the covariate.

In the case of the regression coefficient for X_{1i}, we know that half the difference between the mean Y_i score under the new and old curricula equals 3.25, i.e., the regression coefficient for X_{1i} not controlling for the pretest. The regression coefficient for X_{1i} controlling for the pretest equals 1.04. This equals half the difference in Y_i associated with the difference in curriculum over and above any pretest differences associated with curriculum.

An equivalent but slightly different way to think about the coefficient for a contrast-coded predictor in the presence of a covariate is that it tells us about the magnitude of differences among adjusted values of $\bar{Y}_k$, adjusting those condition means to get rid of differences in the covariate condition means. More precisely, we can compute for each condition the adjusted mean:

$$\bar{Y}'_k = \bar{Y}_k - b_z(\bar{Z}_k - \bar{Z})$$

where $\bar{Z}$ is the mean of the condition means for the covariate. We can then use these adjusted means to derive the regression coefficient for a contrast-coded predictor in the model that includes the covariate, using the old formula for the regression coefficient for a contrast-coded predictor. In other words, once we have the adjusted means, the regression coefficient for a contrast-coded predictor in the model that includes the partially redundant covariate equals:

$$\frac{\sum_k \lambda_k \bar{Y}'_k}{\sum_k \lambda_k^2}$$

where $\bar{Y}_k$– are the adjusted cell means as just defined.

To illustrate, in Figure 10.10 the values of these adjusted cell means, $\bar{Y}'_k$, are given for the four teacher × curriculum conditions of our design. These were derived using the formula for the adjusted cell means above. For example, the value of the adjusted mean for the Old Curriculum × Teacher A condition is given by:

$$\bar{Y}'_k = 56 - 1.26(49 - 50.25) = 57.58$$

FIGURE 10.10 Post-test condition means adjusted for the pretest ($\bar{Y}'_k$)

Condition	$\bar{Y}'_k$
Old Curriculum, Teacher A	57.58
Old Curriculum, Teacher B	57.85
New Curriculum, Teacher A	59.05
New Curriculum, Teacher B	60.52

where 56 is the value of $\bar{Y}_k$ for this condition, 1.26 is the regression coefficient for the pretest in the model that includes both the pretest and the condition contrast-coded predictors, 49 equals the pretest mean for this condition, $\bar{Z}_k$, and 50.25 equals the mean of the four pretest condition means.

We can now use these adjusted $\bar{Y}_k$ to compute the regression coefficients for the contrast-coded predictors. For instance, the coefficient for X_{1i} in the model that includes the covariate equals:

$$\frac{\sum_k \lambda_k \bar{Y}'_k}{\sum_k \lambda_k^2} = \frac{(-1)57.58 + (-1)57.85 + (+1)59.05 + (+1)60.52}{4} = 1.04$$

Conceptually, then, the coefficients for contrast-coded predictors in the presence of a covariate tell us about the magnitude of the coded differences among adjusted condition means, adjusting those dependent variable means by the extent to which the covariate means depart from each other. Obviously, if all $\bar{Z}_k$ are equal to each other there will be no adjustment, as we saw in the case of an orthogonal covariate. But with a partially redundant covariate, the differences among the $\bar{Z}_k$ will result in some adjustment among the dependent variable means that are compared when examining the regression coefficient for a contrast-coded predictor.

The source table for the ANCOVA model with these modified data is given in Figure 10.11. Notice that the sums of squares and, as a result, the F statistics for the omnibus test of condition differences and the individual contrast tests are all considerably smaller than they were in the ANOVA source table. This is so because they are testing different null hypotheses than they were in the ANOVA model. In the ANOVA model the omnibus test was testing whether there were any differences among the condition means $\bar{Y}_k$. The tests of the contrasts were testing specific comparisons among these condition means. With the inclusion of the pretest in the model, the omnibus test is now testing for the presence of differences among the adjusted condition means. Similarly, the contrast tests are now testing specific comparisons among these adjusted condition means. In other words, the tests are now examining condition differences in $\bar{Y}_k$ having adjusted for condition differences that existed on the pretest.

Since the pretest is now partially redundant with the various contrast codes that code condition, the sums of squares in this source table are not additive as they were in the earlier ANOVA table. Therefore, some of them must be derived through the estimation of various compact models. To get the sum of squares for the omnibus condition test, we must estimate a model with only the pretest used as a predictor, since this omnibus

FIGURE 10.11 ANCOVA source table

Source	*b*	*SS*	*df*	*MS*	*F*	*PRE*
Model		869.43	4	217.36	21.87	.71
Pretest	1.26	361.93	1	361.93	36.39	.51
Between conditions		34.30	3	11.43	1.15	.09
Curriculum	1.04	27.81	1	27.81	2.80	.07
Teacher	0.43	7.44	1	7.44	0.75	.02
Curriculum x Teacher	0.30	3.29	1	3.29	0.33	.01
Error		348.07	35	9.94		
Total		1217.50	39			

sum of squares is no longer equal to the sum of the three sums of squares explained by the three contrast-coded predictors. Even though the contrast codes are still orthogonal and even though there are still an equal number of observations in each condition, the three contrast-coded predictors differ in how redundant or correlated they are with the covariate. Thus, the sum of their three individual sums of squares is not equal to the difference in the sum of squares explained if they are all omitted from the model.

Models with two or more partially redundant covariates are simple extensions to the single covariate case that we have just examined. The formula for the adjusted means, for instance, simply adjusts for or subtracts off mean differences on each covariate, each weighted by its regression coefficient. For instance, if Z_{1i} and Z_{2i} are two covariates, then the regression coefficient for contrast-coded predictors in the model that included these two covariates would be examining comparisons among the following (doubly) adjusted cell means:

$$\bar{Y}'_k = \bar{Y}_k - b_{z_1}(\bar{Z}_{1k} - \bar{Z}_1) - b_{z_2}(\bar{Z}_{2k} - \bar{Z}_2)$$

As a final comment, reiterating what we said before, while the ANCOVA permits comparisons among condition means adjusting for one or more partially redundant covariates, this is not a general solution to problems of internal validity in research designs where random assignment of observations to conditions has not been used. Adjustment for redundant covariates is a powerful procedure, but it does not solve the problems that limit causal inferences in nonexperimental research designs.

Examining Partial Effects: The Case of Mediation

All of what we have just said about the analysis of categorical predictor variables when we control for a partially redundant covariate applies as well to the second occasion when such models are of interest, namely, when we wish to examine the process that mediates the effects of some experimental categorical variable on a dependent variable. Let us consider a new example to illustrate the interpretations that ensue from such an analysis in this case. Suppose that participants were randomly assigned to a family counseling intervention designed to improve outcomes for adolescents who suffer from bipolar disorder. Either they receive the family counseling intervention along with the usual pharmacological care or they receive only the usual care. And the question is whether this intervention affects the manifestation of bipolar symptoms eight weeks later. The dependent variable is assessed by a clinical psychologist who is blind to experimental treatment, using a 10-point rating scale (higher scores equal more symptoms).

FIGURE 10.12 Hypothetical data for 20 families

	Treatment		*Control*	
	Criticism (C_i)	*Symptoms (S_i)*	*Criticism (C_i)*	*Symptoms (S_i)*
	3	4	4	7
	2	3	4	8
	4	4	3	5
	2	3	2	5
	2	4	4	7
	3	5	4	6
	1	3	5	6
	2	5	3	4
	3	7	4	5
	3	5	4	6
Mean	2.5	4.3	3.7	5.9

If the family counseling intervention is effective in reducing symptoms, it is thought that it must operate by affecting the amount of criticism the parents direct at the bipolar adolescent during the interim period. Hence, at week 7 of the study, another clinical psychologist, again blind to experimental treatment, interviews the parents and rates the degree to which the parents spontaneously criticize the adolescent, this time on a five-point scale (higher scores equal more criticism). Hence, for each of 20 families there are three variables: whether they received the experimental treatment or not, assessed parental criticism at 7 weeks, and adolescent symptoms at 8 weeks. Hypothetical scores on these variables for the families are given in Figure 10.12.

Obviously one first wishes to assess whether the treatment had an effect on the ultimate outcome variable: symptoms at 8 weeks. If it did, then the theoretical expectation is that those effects are mediated through the parental criticism variable, measured at 7 weeks. In other words, the thinking is that the treatment effect ought to work, if it does, by affecting the manifestation of parental criticism of the adolescent and this in turn is responsible for lower symptom levels.

These expectations imply the following:

1. There will be an overall treatment effect on symptoms at 8 weeks; the treatment will reduce the number of symptoms.
2. The treatment will lead to lower parental criticism.
3. Parental criticism will predict fewer symptoms at 8 weeks holding constant treatment condition.
4. The effect of the treatment on symptoms will be reduced when parental criticism is controlled.

The first two of these would be assessed by a separate two-group ANOVA model for each of the dependent variables, first symptoms at 8 weeks and then parental criticism at 7 weeks. The third and fourth conditions would be assessed by a single ANCOVA model, using both treatment and parental criticism as predictors of symptoms. Here, we expect that criticism will predict fewer symptoms and that the effect of treatment on symptoms will be reduced compared to its effect in the simple ANOVA model. These are the classic conditions for establishing mediation, as identified by Judd and Kenny (1981b) and Baron and Kenny (1986).

We first contrast-code the treatment variable (X_i = +.5 if treatment; –.5 if control) and then examine its effects on both symptoms and criticism with two simple regression models, examining the first two of the above four expectations. The overall treatment effect on symptoms is estimated as:

$$\hat{S}_i = 5.10 - 1.60X_i$$

This model, predicting the two group means, has a sum of squared errors of 27.00. The estimated slope for X_i equals the difference between the two treatment means, given the codes we have used. And a test of whether it departs from zero, and thus whether the two group means on symptoms differ, yields a PRE of .32 and an F of 8.53 with 1 and 18 degrees of freedom. Hence, there is an overall treatment effect; the treatment resulted in fewer symptoms measured 8 weeks later.

The effect of treatment on criticism at 7 weeks is examined in the following simple regression model:

$$\hat{C}_i = 3.10 - 1.20X_i$$

with a sum of squared errors of 12.60. Again, the slope for the treatment variable equals the mean difference between the treatment and control groups in criticism and this difference is significant: PRE = .36, $F_{1,18}$ = 10.29.

To assess the third and fourth mediation expectations, the ANCOVA model is estimated as:

$$\hat{S}_i = 2.84 - 0.72X_i + 0.73C_i$$

with a sum of squared errors of 20.28. In the previous paragraph we showed that the means for criticism differ significantly between the two conditions, demonstrating that the two predictors in this ANCOVA model are redundant with each other. Accordingly, the coefficient for treatment no longer estimates the difference between the two condition means; rather, it estimates the difference between the two conditions when adjusting or controlling for differences between them on the criticism variable:

$$b_{SX.C} = \frac{\sum_k \lambda_k \bar{S}_k}{\sum_k \lambda_k^2} - b_{SC.X} \frac{\sum_k \lambda_k \bar{C}_k}{\sum_k \lambda_k^2} = -1.60 - (0.73)(-1.20) = -0.72$$

Notice in this expression that we are indicating each parameter estimate in this model with multiple subscripts, the first letter indicating the dependent variable, the second the predictor involved, and then, following the dot, any other variables controlled for in the model. The reason for this complete notation will become apparent shortly.

In this model, the test of the slope associated with the covariate, criticism, is significant (PRE = .25, $F_{1,17}$ = 5.63), while that for the treatment variable no longer is (PRE = .08, $F_{1,17}$ = 1.40). Accordingly, it would appear that the four conditions for establishing mediation (outlined above) have been met. Namely, the treatment has an overall effect on the ultimate outcome variable as well as on the mediator. The mediator significantly affects the outcome when controlling for treatment, whereas when controlling for the mediator the effect of treatment on symptoms is no longer significant.

However, such a demonstration does not necessarily mean that there has been a significant reduction in the treatment effect once criticism is controlled. In other words, although the treatment effect no longer differs from zero, it might not significantly differ from its effect in the initial ANOVA model. For this and other reasons, tests of whether the partial regression coefficient for the treatment in the ANCOVA model is significantly smaller than its overall or total effect have been developed. Rather than develop those tests here, we refer the interested reader to MacKinnon, Lockwood, Hoffman, West, and Sheets (2002), MacKinnon (2008), and Judd, Yzerbyt, and Muller (2015). Importantly, however, they rest on the following equality:

$$b_{SX} - b_{SX.C} = b_{SC.X} b_{CX}$$

which means that the reduction of the overall effect of the treatment when the mediator is controlled equals the product of the treatment's effect on the mediator and the mediator's partial effect on the outcome controlling for the treatment.

This equality is none other than our formula for the partial slope associated with a contrast-coded predictor when a redundant covariate is controlled, as just given:

$$b_{SX.C} = \frac{\sum_k \lambda_k \bar{S}_k}{\sum_k \lambda_k^2} - b_{SC.X} \frac{\sum_k \lambda_k \bar{C}_k}{\sum_k \lambda_k^2} = b_{SX} - b_{SC.X} b_{CX}$$

A small amount of algebraic manipulation gives the desired equality: $b_{SX} - b_{SX.C} = b_{SC.X} b_{CX}$. In other words, the tests that have been developed to determine whether the treatment effect is significantly reduced when the mediator is controlled are tests of the partial regression slope associated with the treatment variable. If such tests are significant (i.e., the regression slope is significantly smaller in the ANCOVA model), the treatment variable is said to have an indirect effect on the outcome variable via the mediator.

This mediation example is obviously rather simple and our treatment does not do justice to the extensive literature on the subject. Our point has been simply to illustrate that everything we have said about ANCOVA with a partially redundant covariate applies in the case of mediation assessment as well as in the case of simply a confounded treatment variable. These two differ not in the analyses conducted but in their underlying theoretical model.

The Homogeneity of Regression Assumption in ANCOVA

Most classic treatments of the ANCOVA specify that an assumption, referred to as the homogeneity of regression assumption, is crucial to the use and interpretation of ANCOVA results. To clarify this assumption, why it is potentially important, and what may be done if it is violated, let us return to the data example used earlier, with the two crossed independent variables being curriculum (new versus old) and teacher (A versus B) and a confounded covariate, the pretest. The data we will use were given in Figure 10.8 and the cell means, for both the post-test and the pretest, were given in Figure 10.9. We use the same notation we did earlier: the post-test is Y_i, the pretest is Z_i, X_{1i} is the contrast-coded predictor representing curriculum (+1 if new; –1 if old), X_{2i} is the contrast-coded predictor representing teacher (–1 if A; +1 if B), and X_{3i} is their product, representing the curriculum × teacher interaction.

The ANCOVA model that we presented earlier for these data was estimated as:

$$\hat{Y}_i = -4.84 + 1.26Z_i + 1.04X_{1i} + 0.43X_{2i} + 0.30X_{3i}$$

$$SSE = 348.07$$

And we saw that the slopes associated with the contrast-coded predictors could be interpreted either as differences among Y_i cell means, adjusting for parallel differences among the Z_i cell means:

$$\frac{\sum_k \lambda_k \bar{Y}_k}{\sum_k \lambda_k^2} - b_Z \frac{\sum_k \lambda_k \bar{Z}_k}{\sum_k \lambda_k^2}$$

or as differences among the adjusted Y_i cell means, where those adjusted means are given by the expression:

$$\bar{Y}'_k = \bar{Y}_k - b_Z(\bar{Z}_k - \bar{Z})$$

The assumption of homogeneity of regression that underlies this analysis, and the interpretation of its parameter estimates, is that the relationship between the pretest and the post-test is invariant across the four conditions defined by the contrast-coded predictors. To understand this, note that in these expressions for the interpretation of the slopes associated with the contrast-coded predictors we use one value for the slope of the pretest, b_Z, rather than different values for the four different conditions. That is, when calculating the adjusted cell means that are compared, a single value of b_Z is used regardless of the value of k. If the relationship between the dependent variable, Y_i, and the covariate, Z_i, differs substantially in magnitude across the various conditions, then we should not be assuming a single adjustment weight b_Z, but instead should be allowing for different adjustment weights for the various conditions. Accordingly, in controlling for a covariate and interpreting the resulting parameter estimates for the contrast-coded predictors, it makes sense to examine whether the relationship between Y_i and Z_i is invariant or homogeneous across conditions.

To suggest that the relationship between the pretest and the post-test depends on condition is to suggest that condition and the pretest interact in affecting the post-test. Therefore, to test the homogeneity of the relationship between Y_i and Z_i across conditions, we need to test whether the interactions between Z_i, on the one hand, and the condition defining contrast-coded predictors, on the other, are significant. To examine these interactions, we follow the standard procedure of computing products of the variables whose interactions we wish to test and then entering those product variables as separate predictor variables into a model that includes the variables that are components of the products. We then test whether the augmented model that includes the product terms generates significantly better predictions than a compact model that omits all of the pretest by contrast-coded predictor interactions.

Let us illustrate this using the data contained in Figure 10.8. Since we have three contrast codes that define condition, there will be three interaction or product terms to examine the condition by pretest interaction: $X_{1i}Z_i$, $X_{2i}Z_i$, and $X_{3i}Z_i$. When these are included in the model, the parameter associated with the first will estimate the extent to which the pretest–post-test relationship is homogeneous across the two curriculum levels, the second will estimate the extent to which it is homogeneous across the two teachers, and the third will examine the triple interaction, i.e., whether the pretest–post-test relationship is homogenous across levels of the curriculum × teacher interaction.

The model that includes all three of these pretest × condition interactions is estimated as:

$$\hat{Y}_i = -4.40 + 1.26Z_i + 6.11X_{1i} - 1.32X_{2i} - 3.93X_{3i} - 0.10X_{1i}Z_i + 0.03X_{2i}Z_i + 0.08X_{3i}Z_i$$

$$\text{SSE} = 344.23$$

To demonstrate that this model in fact does allow the slope for the covariate to vary between the four cells of our design, let us examine the expression for the "simple" effect of the covariate in the various cells. The model can be rewritten as:

$$\hat{Y}_i = (-4.40 + 6.11X_{1i} - 1.32X_{2i} - 3.93X_{3i}) + (1.26 - 0.10X_{1i} + 0.03X_{2i} + 0.08X_{3i})\, Z_i$$

And from this, we can derive the "simple" $Y_i : Z_i$ regression models in each of the four cells of the design, substituting for the various values of X_{1i}, X_{2i}, and X_{3i}. For instance, for the New Curriculum, Teacher B cell, we get:

$$\hat{Y}_i = (-4.40 + 6.11(+1) - 1.32(+1) - 3.93(+1)) + (1.26 - 0.10(+1) + 0.03(+1) + 0.08(+1))\, Z_i$$

which reduces to:

$$\hat{Y}_i = -3.54 + 1.27Z_i$$

Parallel expressions for the other three cells of the design give:

New Curriculum, Teacher A: $\hat{Y}_i = 6.96 + 1.05Z_i$

Old Curriculum, Teacher B: $\hat{Y}_i = -7.90 + 1.31Z_i$

Old Curriculum, Teacher A: $\hat{Y}_i = -13.12 + 1.41Z_i$

Clearly this model allows different "simple" slopes for Z_i in the four cells of the design.

Since we wish to examine whether there are any differences among conditions in these "simple" slopes, and since we have no specific expectations that predict such differences, we can conduct an omnibus three-degree-of-freedom test to examine the homogeneity of regression assumption, testing whether the set of three interaction terms leads to a significant improvement in the fit of the model. Thus this interactive model becomes Model A and the earlier ANCOVA model is the Model C with which we want to compare it. This comparison yields the following PRE and F values:

$$\text{PRE} = \frac{348.07 - 344.23}{348.07} = .011$$

$$F_{3,32} = \frac{.011/3}{(1 - .011)/32} = \frac{3.845/3}{344.226/32} = 0.119$$

There is no reason to prefer Model A over Model C. Thus there is no evidence in these data that the homogeneity of regression assumption of ANCOVA has been violated. Equivalently, we have found no evidence to suggest that the relationship between the pretest and the post-test differs depending on which of the four cells of the research design we are looking at. Using a single slope, and thus a single adjustment weight, to compute the four adjusted means suffices.

As a check on the homogeneity of regression assumption, this omnibus model comparison seems sufficient. However, there may be times when one has prior expectations that one or more of the interactions between the pretest and the contrast-coded predictors would be found. For instance, it is possible that one might have expected the pretest–post-test relationship to be stronger under the old curriculum than under the new one. In such a case, it seems appropriate to examine the slope associated with that particular interaction term, $X_{1i}Z_i$, comparing the augmented model that includes all three interaction terms to a compact one that only leaves out this one hypothesized interaction.

Should it be found that the homogeneity of regression assumption has been violated, either through the omnibus test of the set of interaction terms, or the more focused examination of single interaction terms in the case of prior expectations about their possible existence, then one is left with the model that includes one or more covariate × condition interactions and one is forced to interpret them. Importantly, then, the homogeneity of regression assumption is an assumption whose violation implies simply that a model more complicated than the ANCOVA model is required and one must interpret the obtained significant interaction term(s). Thus, this assumption is not like the assumptions about the distribution of residuals that we made very early on in this book (i.e., that they are normally distributed, have a single variance, and are independent). Violations of those assumptions mean that our inferential tests are biased. In this case, violations of the homogeneity of regression assumption simply mean that life is a bit more complicated and one must interpret the data in light of the resulting covariate × condition interactions.

Suppose, for instance, in the present case, that the interaction between Z_i and X_{1i} had in fact proven to be significant. One would then be compelled to provide interpretations in the context of the more complicated interactive model that included at least this significant interaction. And in this case, the resulting interaction could be interpreted by focusing either on the extent to which the pretest–post-test relationship depends on the new versus old curriculum or on the extent to which the magnitude of the new – old curriculum difference depends on the value of the pretest. Although these two interpretations are fundamentally equivalent, given our focus on the condition differences the second is likely to be the preferred interpretation of the interaction coefficient. We will proceed to illustrate the interpretation with the present data, even though our omnibus test clearly failed to show any evidence of the interactions, and the more focused single-degree-of-freedom test of just $X_{1i}Z_i$ is also not significant in these data.

As we saw in Chapter 7, the interpretation of parameter estimates in the presence of significant interactions involving continuous predictors is complicated by the fact that slopes of variables that are components of included product predictors estimate "simple" effects at the value of zero of the other component variables included in the product predictor. Thus, in the case of the current Model A, the estimated slope for X_{1i} (i.e., 6.11) informs us about half the expected curriculum difference for someone scoring zero on the pretest variable, Z_i. Since there are no pretest values in our dataset that are close to zero, interpretations of the parameter estimates in the Model A are rendered considerably more meaningful if one centers or mean-deviates the covariate in the model that includes the products of the contrast-coded predictors with that centered covariate. The resulting Model A in this centered case is:

$$\hat{Y}_i = 58.86 + 1.26Z_{Ci} + 1.00X_{1i} + 0.36X_{2i} + 0.27X_{3i} - 0.10X_{1i}Z_{Ci} + 0.03X_{2i}Z_{Ci} + 0.08X_{3i}Z_{Ci}$$

where Z_{Ci} is the centered version of the pretest (i.e., $Z_{Ci} = Z_i - \bar{Z}$). In a deep sense, of course, this centered model is identical to the model prior to centering, in that it makes the same predictions and has the same sum of squared errors (i.e., 344.23). But now the slope estimates associated with the three contrast-coded predictors are considerably more interpretable: they estimate the "simple" condition differences at the mean value of the pretest. For instance, the slope associated with X_{1i} in this model (i.e., 1.00) estimates half the predicted difference in post-test scores between the old and new curriculum conditions at the mean value of the pretest and allowing the magnitude of the old – new curriculum difference to depend on the value of the pretest. And the degree to which that difference does depend on the value of the pretest is indicated by the parameter estimate associated with the $X_{1i}Z_{Ci}$ product predictor. As the value of the pretest increases by one unit, the estimated "simple" difference attributable to the old versus new curriculum decreases by 0.10 units. Had this value been significant, we would have concluded that curriculum effects are larger the less well one initially performed.

An equivalent interpretation for this parameter estimate can be given by focusing on the "simple" slopes of the pretest in the two different curriculum conditions derived earlier. That "simple" slope, on average, across the two conditions is 1.26. The parameter estimate associated with the $X_{1i}Z_{Ci}$ product predictor tells us that the "simple" slope is 0.10 smaller in the new curriculum condition and 0.10 units larger in the old curriculum condition. In other words, the pretest–post-test relationship is stronger under the old curriculum than under the new one.

Our interpretation of covariate × condition interactions in this example has been a bit forced because in fact the data suggest that the homogeneity of regression assumption is not violated. Our purpose has been simply to illustrate how one might proceed if it happened that the assumption *was* violated. One would simply live with the more complex model and interpret the resulting significant covariate × condition interactions. But there are other conditions where such interactions are the primary focus of research, where one is precisely interested in whether the magnitude of the relationship between some continuous predictor variable (heretofore called a covariate) and the dependent variable depends on the levels of one or more categorical variables. And it is to such an example that we now turn.

MODELS WITH CONTINUOUS AND CATEGORICAL PREDICTORS OUTSIDE OF EXPERIMENTAL CONTEXTS

Our data come from a large western public university and consist of the academic records of 2740 members of the freshman class, either enrolled in the College of Arts and Sciences or in the College of Engineering. We have four variables available on these students: their combined SAT score, taken during their senior year in high school, on a scale of 20–80 (average of verbal and math SAT scores dropping the final digit); their cumulative

FIGURE 10.13 SAT scores and freshman GPAs by college and gender

	SAT		*GPA*	
	Male	*Female*	*Male*	*Female*
Engineering				
Mean	63.50	61.30	2.88	3.03
SD	(6.02)	(5.60)	(0.70)	(0.63)
n	329	91	329	91
Arts and Sciences				
Mean	58.39	56.62	2.68	2.93
SD	(6.64)	(6.54)	(0.76)	(0.67)
n	968	1352	968	1352

grade point average (GPA) at the end of the freshman year; the college in which they were enrolled; and their gender. We are interested in knowing whether applicants' SAT scores predict their freshman GPA at university and whether that relationship depends both on the college in which they enrolled and gender.

The mean combined SAT scores and freshman GPAs for each of the four groups, defined by college and gender, are given in Figure 10.13

We start by examining mean differences as a function of gender and college in each of these variables, defining X_{1i} as +1 if Female, –1 if Male, X_{2i} as +1 if Arts and Sciences, –1 if Engineering, and X_{3i} as their product. The two ANOVA models, one for each variable, are given below:

$$\widehat{SAT_i} = 59.95 - 0.99X_{1i} - 2.44X_{2i} + 0.11X_{3i} \qquad \text{SSE} = 115{,}139.16$$

$$\widehat{GPA_i} = 2.880 + 0.100X_{1i} - 0.077X_{2i} + 0.025X_{3i} \qquad \text{SSE} = 1355.40$$

For both variables, the gender difference is significant, although in opposite directions. Females have lower combined SAT scores on average than males ($F_{1,2736} = 124.71$, PRE = .04) but they end up with higher freshman year GPAs ($F_{1,2736} = 58.56$, PRE = .02). There is also a significant college difference for both variables, with students in Engineering having both higher SAT scores ($F_{1,2736} = 196.53$, PRE = .07) and higher freshman GPAs ($F_{1,2736} = 21.39$, PRE = .01) than Arts and Sciences students.

A simple regression model in which SAT scores are used to predict freshman GPAs yields the following parameter estimates:

$$\widehat{GPA_i} = 1.018 + 0.031SAT_i \qquad \text{SSE} = 1269.97$$

Unsurprisingly, students with higher combined SAT scores when they enter university end up with better freshman year GPAs ($F_{1,2738} = 270.97$, PRE = .09).

We have already seen that SAT scores are significantly related to the two categorical variables of college and gender. Therefore, when we estimate what we have to this point called the ANCOVA model, with SAT and X_{1i}–X_{3i} as predictors of GPA, these predictors will be partially redundant. The resulting model is:

$$\widehat{GPA_i} = 1.018 + 0.031SAT_i \qquad \text{SSE} = 1269.97$$

FIGURE 10.14 ANCOVA source table

Source	*SS*	*df*	*MS*	*F*	*PRE*
Model	187.16	4	46.79	105.89	.13
Gender (X_1)	18.57	1	18.57	41.99	.02
College (X_2)	0.10	1	0.10	0.22	.00
Gender x College (X_3)	0.46	1	0.46	1.04	.00
SAT	146.91	1	146.91	332.33	.11
Error	1208.49	2735	0.44		
Total	1395.65	2739			

And the source table in Figure 10.14 summarizes the statistical results.

Clearly, SAT remains a significant predictor of freshman GPA when controlling for gender, college, and their interaction. Additionally, of the categorical predictors, only gender remains significant once SAT is controlled. The effect of college, which was significant in the ANOVA model, no longer is. It seems that once students are equated in terms of SAT performance, there is no longer a college difference in GPAs. As with any ANCOVA model, the slopes for the categorical predictors are now informing us about the magnitude of differences among the adjusted GPA cell means. These are computed as:

$$\overline{GPA_k}' = \overline{GPA_k} - b_{SAT}\,(\overline{SAT_k} - \overline{SAT})$$

and their values are:

For male Engineering students:	2.69
For female Engineering students:	2.92
For male Arts and Sciences students:	2.68
For female Arts and Sciences students:	2.99

As we have already discussed, this model assumes that the "simple" slope of SAT does not vary across the four groups. That is, this model makes the homogeneity of regression assumption that we discussed earlier. It is of central interest in these data to ask whether the relationship between SAT and freshman GPA is the same across all four student groups. Perhaps SAT performance is more predictive of GPA in one college or the other. If so, then the greater diagnosticity of the test among the students in one college than the other could, perhaps, be taken into account at the time of admissions, weighting more heavily the SAT scores in making admissions decisions for some students than for others. Similarly, a gender difference in the diagnosticity of the test would indicate predictive bias, that is, that the test makes better predictions for one gender than the other, which would likely need to be addressed in other ways, for example by revising the test itself. In any case, if SAT scores are more diagnostic among some student groups than others, then it implies that the homogeneity of regression assumption would be violated in these data: we would need different SAT slopes for students who differ in their gender, their course of study, or the interaction of these two factors.

To test whether SAT is differentially related to GPA among the four different groups, we estimate a model that includes as predictors all products of SAT with the contrast-coded predictors that code the four categories of students: SAT_iX_{1i}, SAT_iX_{2i}, and SAT_iX_{3i}.

But this time, since we suspect that SAT performance will vary in its diagnosticity across the four groups, we will examine the individual contributions of each of these product predictors rather than simply conducting the omnibus test of whether they, as a set, increase the explanatory power of the model. The estimated Model A is:

$$\widehat{GPA_i} = 0.200 - 0.260X_{1i} + 0.676X_{2i} + 0.289X_{3i} + 0.044SAT_i + 0.007SAT_iX_{1i} - 0.011SAT_iX_{2i} - 0.004SAT_iX_{3i}$$

$$\text{SSE} = 1204.18$$

Tests of the individual product predictors are summarized in the partial source table of Figure 10.15. Although the effects sizes are not large here, there is clearly evidence to suggest that the relationship between SAT and freshman GPA depends on whether one is an Engineering student or an Arts and Sciences student.

FIGURE 10.15 Portion of source table testing interactions between continuous and categorical predictors

Source	*SS*	*df*	*MS*	*F*	*PRE*
SAT x Gender ($SATX_1$)	1.45	1	1.45	3.28	.001
SAT x College ($SATX_2$)	3.92	1	3.92	8.88	.003
SAT x Gender x College ($SATX_3$)	0.64	1	0.64	1.46	.001
Error	1204.18	2732	0.44		

To interpret these differences, let us re-express the model in terms of the "simple" effect of SAT for each of the four groups:

$$\widehat{GPA_i} = (0.200 - 0.260X_{1i} + 0.676X_{2i} + 0.289X_{3i}) + (0.044 + 0.007X_{1i} - 0.011X_{2i} - 0.004X_{3i})SAT_i$$

This re-expression can be used to generate the "simple" GPA : SAT prediction functions for each of the four cells by substituting the appropriate values for the contrast-coded predictors.

For male Engineering students:

$$\begin{aligned}\widehat{GPA_i} &= (0.200 - 0.260(-1) + 0.676(-1) + 0.289(+1)) + (0.044 + 0.007(-1) - 0.011(-1) - 0.004(+1))SAT_i \\ &= 0.073 + 0.044SAT_i\end{aligned}$$

For female Engineering students:

$$\begin{aligned}\widehat{GPA_i} &= (0.200 - 0.260(+1) + 0.676(-1) + 0.289(-1)) + (0.044 + 0.007(+1) - 0.011(-1) - 0.004(-1))SAT_i \\ &= -1.025 + 0.066SAT_i\end{aligned}$$

For male Arts and Sciences students:

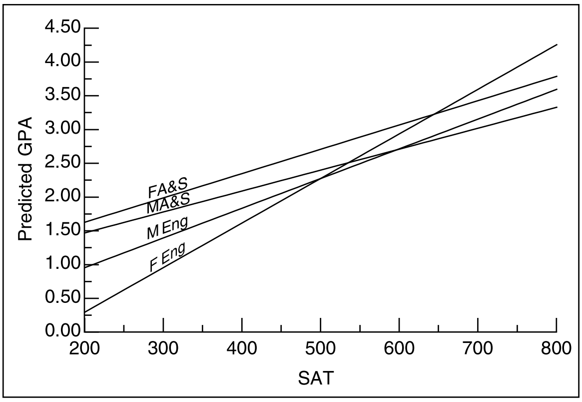
$$\widehat{GPA_i} = (0.200 - 0.260(-1) + 0.676(+1) + 0.289(-1)) + (0.044 + 0.007(-1) - 0.011(+1) - 0.004(-1))SAT_i$$
$$= 0.847 + 0.031SAT_i$$

For female Arts and Sciences students:

$$\widehat{GPA_i} = (0.200 - 0.260(+1) + 0.676(+1) + 0.289(+1)) + (0.044 + 0.007(+1) - 0.011(+1) - 0.004(+1))SAT_i$$
$$= 0.906 + 0.036SAT_i$$

In Figure 10.16 we have graphed these four "simple" relationships. As the differences in these graphed slopes make clear, the two Engineering groups have steeper slopes than do the two Arts and Sciences groups. This is the implication of the significant SAT × college interaction, with its coefficient, –0.011, equaling half the difference between the average of the two "simple" slopes for the Arts and Sciences student groups and the average of the two "simple" slopes for the Engineering student groups. In other words, when averaging across the two gender groups, SAT performance is more diagnostic of freshman year GPA among Engineering students than it is among Arts and Sciences students.

FIGURE 10.16 Four-group simple relationships between SAT and GPA



A result that is perhaps somewhat surprising is that if we were to conduct four simple regressions, one for each of the four student groups and regressing freshman GPA on SAT scores for just the students in those four groups, we would get exactly the above parameter estimates that we have just calculated as the "simple" coefficients from this model that includes the SAT × interaction terms. That is, we split our sample into the four groups, and in each group separately we regress GPA on SAT, with the following results:

For male Engineering students:

$$\widehat{GPA}_i = 0.073 + 0.044SAT_i \qquad \text{SSE} = 135.75$$

For female Engineering students:

$$\widehat{GPA}_i = -1.025 + 0.066SAT_i \qquad \text{SSE} = 23.77$$

For male Arts and Sciences students:

$$\widehat{GPA}_i = 0.847 + 0.031SAT_i \qquad \text{SSE} = 520.25$$

For female Arts and Sciences students:

$$\widehat{GPA}_i = 0.906 + 0.036SAT_i \qquad \text{SSE} = 524.41$$

In a deep sense, then, the full model that includes all three contrast-coded predictors and their interactions with SAT is equivalent to four simple regression models, one from each of the four student groups, regressing GPA on SAT. Accordingly, if we add up the four sums of squared errors from these four within-group simple regression models, we get the sum of squared errors from the overall interactive model (i.e., 135.75 + 23.77 + 520.25 + 524.41 = 1204.18).

While each of the within-group simple regressions provides us with a test of whether there is a significant association between SAT and GPA in that group (and all four are significant in this case), it is only by testing the various interaction terms in the full model, using all the data, that we are able to examine statistically whether the various simple slopes differ from each other across the four groups. And in this case, while all four simple slopes are significant, the two for the Engineering students are significantly larger than those from the Arts and Sciences groups.

In general, then, to examine whether two variables are related more strongly to each other in some groups than they are in others, we would encourage a test of whether the group $\times$ continuous predictor interactions are significant in a model using the data from all groups together. This might be accompanied by separate model estimates in each of the groups, to test whether the group-specific simple slopes differ from zero. But to determine whether these simple slopes are different from each other, the test of the interactions in the full model, with all the data, is necessary.

Before leaving this final example, we re-estimate the full interactive model, this time centering or mean-deviating SAT (represented as SAT_{Ci}), both as a predictor and in computing the product predictors:

$$\begin{aligned}\widehat{GPA}_i = {} & 2.784 + 0.124X_{1i} + 0.045X_{2i} + 0.033X_{3i} + 0.044SAT_{Ci} + 0.007SAT_{Ci}X_{1i} \\ & - 0.011SAT_{Ci}X_{2i} - 0.004SAT_{Ci}X_3 \\ \text{SSE} = {} & 1204.18\end{aligned}$$

We do this simply to make the parameter estimates associated with the contrast-coded predictors (X_{1i}–X_{3i}) more interpretable. Recall that the slopes associated with components of product predictors represent “simple” slopes of that component variable when and only when the other component equals zero. In the model with SAT uncentered, the parameter estimates associated with the contrast-coded predictors thus inform us about "simple" differences among the students in the four cells of the design (defined by gender × college) when and only when SAT = 0, which is an impossible value. To render these coefficients more interpretable, we therefore center the SAT variable. Thus, in this re-estimated model, the slope associated with X_{1i} estimates the gender difference in freshman GPA for students whose SAT scores are at the average for the sample. In a deep sense, of course, this model is identical to the one estimated prior to centering. Note, additionally, that we do not need to center the contrast-coded predictors. Although their means are not equal to zero (given the very unequal n values of the cells), the zero values do represent the mean of the gender and college categories.

SUMMARY

In many ways, this chapter marks the final point in the process of developing more complex models for our data in order to ask more complex, and perhaps more interesting, questions of those data. Throughout the development of these more complex models we have kept the same basic machinery to determine whether the increase in complexity is worthwhile as more parameters are added to the model. This machinery depends on comparisons of augmented and compact models in a manner that, by now, ought to be totally routine. While this machinery has remained constant across the chapters, our models have developed from the simplest one involving a single parameter (Chapter 4) to ones making predictions conditional on a single continuous variable (Chapters 5) to ones involving multiple continuous predictors, including product terms (Chapters 6 and 7), to models with categorical variables including products of those categorical variables (Chapters 8 and 9), and finally to models involving continuous and categorical predictor variables and their products (this chapter). This is as far as we wish to extend the complexity of the models we have considered. We believe that nearly every interesting substantive question that social science researchers might like to ask of their data can be answered by using the range of models that have been explored. This is not to say that the limits of model complexity have been reached. Rather, this is to suggest that the models we have considered are those that are most likely to be of use to the data analyst. Further, when other more complicated models seem appropriate, we hope that the reader by now is equipped to adapt the model comparison approach that we have used to these other more complicated situations.

So what remains to be done? The remaining chapters of the book are devoted to problems that are frequently encountered in data that violate various assumptions underlying the model comparison approach we have developed. In the next two chapters, we consider how our models and data analyses need to be modified when dealing with data for which we cannot assume independence of errors. In the final chapters, we focus on violations of the assumptions that residuals are normally distributed and come from a single population, having a single variance.