9

Interactions With Categorical Variables

9.1 NOMINAL SCALE BY NOMINAL SCALE INTERACTIONS

In the last chapter we introduced four coding systems—dummy codes, unweighted effect codes, weighted effect codes, and contrast codes—that may be used to represent nominal (categorical) IVs. We also saw that regression equations could easily be specified that contain a mixture of nominal and quantitative IVs. However, the presentation in Chapter 8 was limited in two important ways. First, we did not consider any models that contained interactions between IVs. And second, we did not consider models in which the effect of the quantitative IV was nonlinear.

The purpose of this chapter is to address these more complex regression models that contain nominal IVs. We begin by considering a series of regression models that contain interactions between nominal variables. The progression of these models starts with the simplest case of the 2 by 2 design and then considers regression models of increasing complexity up to one that includes the interaction between three nominal variables. We then turn to regression models involving the interaction of quantitative and nominal IVs. Finally, we consider regression models that specify nonlinear interactions between nominal and quantitative variables. The material in this chapter builds not only on the foundation for the treatment of nominal IVs presented in Chapter 8, but also on the presentations in earlier chapters on the treatment of nonlinear effects (Chapter 6) and interactions (Chapter 7) of quantitative IVs in MRC.

9.1.1 The 2 by 2 Design

To provide a concrete illustration for our discussion of the 2 by 2 design, we present an example that we will use in the first part of this chapter. Imagine an experiment in which Y is the number of performance errors on a standard task made by each of a sample of rats. One factor in the design is surgery condition, in which the rats are divided into those receiving surgery that destroys a portion of the frontal lobes of their brains (the frontal group); and those whose surgery results in no brain destruction (the sham group). The second factor is drug condition: Within each of the two surgery groups, some rats receive an active drug (the active group), which is expected to minimize the effect of the frontal lobe destruction, and the remainder are given a placebo (the placebo group). Combining the two experimental factors, there are four treatment conditions:

Condition 1: Frontal, active
Condition 2: Sham, active
Condition 3: Frontal, placebo
Condition 4: Sham, placebo

The researcher is interested in answering three questions. First, is there an overall effect of the surgery condition: Do the rats receiving frontal lobe damage (frontal) make more errors on average than those without brain damage (sham)? Second, is there an overall effect of the active drug: Do the rats on active drug make fewer errors on average than rats receiving the placebo? Third, is the magnitude of the effect of frontal damage conditional on the drug condition? This third question asks whether there is a statistical interaction between the surgery and drug effects.

As we saw in Chapter 7, two variables are said to interact in their accounting for variance in Y when *over and above* any additive combination of their separate effects, they have a *joint* effect. In the case of two nominal IVs, novices often confuse the simple linear combination of two main effects with their joint effect, which is something quite different. Table 9.1.1 gives three 2 by 2 tables of means designed to illustrate this distinction. These data are hypothetical and are used to illustrate three different potential outcomes.

In each of the panels of Fig. 9.1.1 and as shown in Table 9.1.1, there is a nonzero "main" effect of both drug and surgery. A main effect is more properly referred to as an *average* effect, that is, it is the effect of a factor averaged over the levels of the other factor. In Table 9.1.1A, for example, rats in the active drug group have made an average effect of four fewer errors than those in the placebo group ($5 - 9 = -4$), and those with frontal lesion had an average of two more errors than those with the sham lesion ($8 - 6 = 2$). However, in the special case considered in Table 9.1.1A each of these average effects is *constant* over the levels of the other, for example, the effect of surgery under the active condition, $6 - 4 = 2$, is exactly the same as

TABLE 9.1.1
Y Means Illustrating No Interaction, Crossed Interaction, and Ordinal Interaction[a]

CH09EX01

	A. No interaction			B. Crossed interaction			C. Ordinal interaction		
	Frontal	Sham	Mean of means	Frontal	Sham	Mean of means	Frontal	Sham	Mean of means
Active	6	4	5	14	4	9	8	2	5
Placebo	10	8	9	10	12	11	12	10	11
Mean of means	8	6	7	12	8	10	10	6	8

	A. No interaction	B. Crossed interaction	C. Ordinal interaction
$M_A - M_P = B_D$	$5 - 9 = -4$	$9 - 11 = -2$	$5 - 11 = -6$
$M_F - M_S = B_S$	$8 - 6 = 2$	$12 - 8 = 4$	$10 - 6 = 4$
$M_{AF} - M_{AS} = B_{(S)A}$	$6 - 4 -= 2$	$14 - 4 = 10$	$8 - 2 = 6$
$M_{PF} - M_{PS} = B_{(S)P}$	$10 - 8 = 2$	$10 - 12 = -2$	$12 - 10 = 2$
$B_{(S)A} - B_{(S)P} = B_{S \times D}$	$2 - 2 = 0$	$10 - (-2) = 12$	$6 - 2 = 4$
B_0	7	10	8

[a]For simplicity, means are rounded to the nearest integer. For generality, the numbers in the cells all differ.

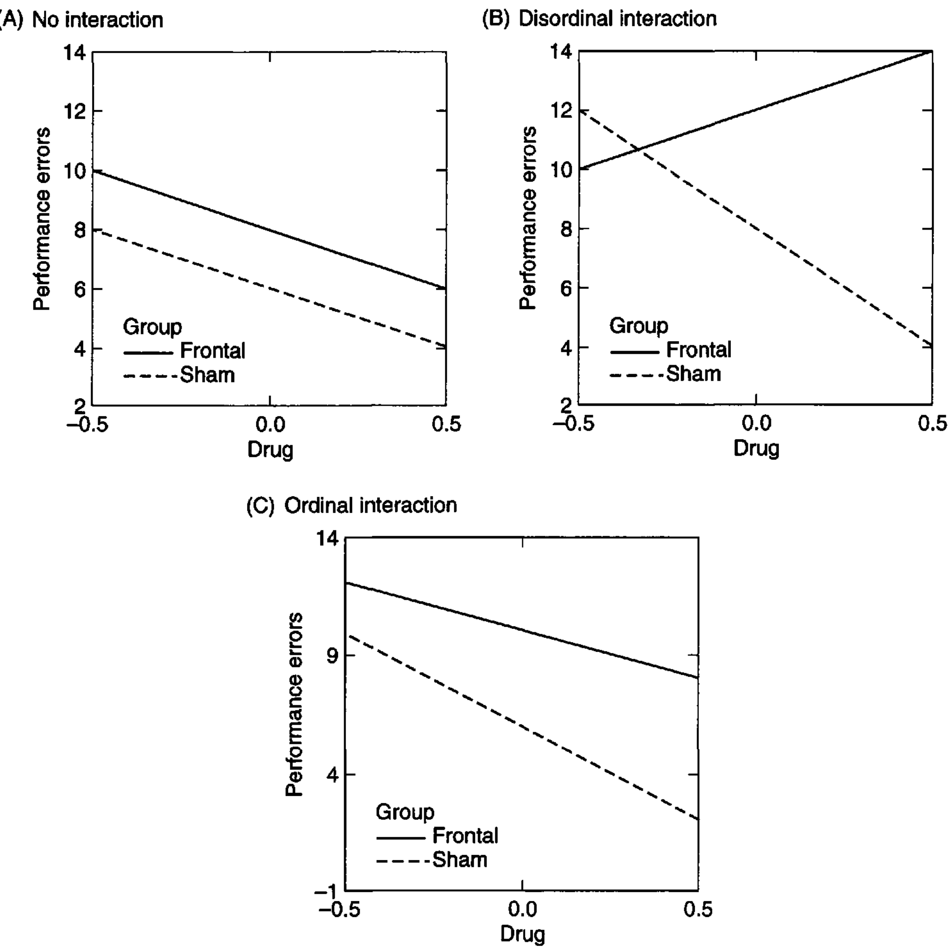

FIGURE 9.1.1 Types of interactions with nominal variables.

its effect under the placebo condition, $10 - 8 = 2$. Similarly, and necessarily, because there is only one *df* for the interaction in a 2 by 2 table, the average (main) effect for active versus placebo of -4 holds for both the frontal ($6 - 10 = -4$) and the sham groups ($4 - 8 = -4$). No interaction exists—each effect, whatever it may be, operates quite independently of the other or, equivalently, quite uniformly for each level of the other.

Consider, by way of contrast, Table 9.1.1B which presents a crossed (or disordinal) interaction. Here, the average mean for the frontal groups is 4 more errors than for sham, $12 - 8 = 4$. However, this is not uniform: for the animals given the active drug, the frontal cases average 10 *more* errors than sham ($14 - 4 = 10$), whereas for those receiving a placebo the frontal cases average 2 *fewer* errors ($10 - 12 = -2$). These two separate effects average out to the surgery main effect $(10 - 2)/2 = 4$, but are obviously quite different, and the fact of their differences constitutes the interaction. The interaction is said to be crossed or disordinal because the effects are of opposite sign ($+10$ and -2). Necessarily it is also crossed if one examines the separate drug effects for frontal ($14 - 10 = 4$) and for control ($4 - 12 = -8$) lesions.

The oppositeness of the signs of the differences is what makes the interaction crossed, but what reveals the interaction is the fact that the effects are different in magnitude. Table 9.1.1C illustrates an ordinal interaction where lines joining the means do not cross within the observed range of the data.[1] The difference between the means of the frontal and sham conditions for animals in the active condition is $8 - 2 = 6$, and for placebo $12 - 10 = 2$. These differences are of the same sign, hence ordinal. Although for both active and placebo groups the frontal lesion results in more errors, this occurs to different degrees, hence an interaction is present. Again, of course, the interaction is also apparent if one takes differences in means vertically instead of horizontally.

Table 9.1.1 should make clear what is meant by a joint effect. Only in Part A can one account for the means in terms of one constant effect for active drug versus placebo and a second constant effect for frontal versus control. In Parts B and C the two factors have an additional joint effect—the combination active drug-frontal and the combination placebo-control have a larger mean number of errors and the active drug-control and frontal-placebo combinations have a smaller mean number of errors than are accounted for by the average effect of drug and the average effect of surgery.

How would the researcher analyze the data corresponding to these 2 by 2 tables using MRC? As we saw in Chapter 8, $g - 1 = 3$ code variables are required to represent the four treatment conditions (cells). The central issue faced by the researcher would be to choose the best coding system to represent his research questions. As we saw in Chapter 8, there are a number of different choices—dummy codes, unweighted effect codes, weighted effect codes, and contrast codes. Recall that dummy codes contrast the mean of each of the other groups with the mean of a selected reference group, unweighted effect codes make comparisons with the unweighted mean of the groups, weighted effect codes make comparisons with the weighted mean of the groups, and contrast codes make a specific set of planned comparisons.

Earlier in this chapter, we noted that the researcher had three research questions corresponding to the average effect of drug, the average effect of surgery, and the conditionality of the surgery effect on the drug condition, the drug by surgery interaction. This set of hypotheses is best represented by a set of contrast codes. The researcher would code rats in the active drug condition $+.5$ and $-.5$ for those in the placebo condition for the first IV, C_D. The researcher would use $+.5$ for rats in the frontal lesion condition and $-.5$ for those in the sham condition for C_S, the surgery variable. Finally, the interaction is a function of the product of these codes on C_D and C_S, that is, $(C_D \times C_S = C_{S \times D})$.[2] When we analyze these data by MRC using these codes, we obtain the regression coefficients given below each 2 by 2 table. In full accord with what has been noted, the interaction $B_{S \times D}$ is 0 for Table 9.1.1, Part A, but takes on nonzero values for Parts B and C. This outcome reflects the requirement that over and above whatever average effects the two research factors have (reflected in B_D and B_S respectively), a third source of Y variation, namely their joint or interaction effect is operating in the latter two data sets. Note also that with contrast coding B_0 is the mean of all the cell means.

Providing that all of the $g-1$ df are represented as code variables, any of the coding schemes presented in Chapter 8 will produce the same R^2 and significance test for the contribution of the set as a whole to the prediction of Y. Here there are four treatment groups, and any coding scheme that produces three less than perfectly correlated IVs will do. However, when the two dichotomous average effects of the 2 by 2 design and their product are coded by the method of contrast coding, not only R^2 and its F ratio are produced, but also meaningful values of B_i, sr_i, pr_i, the t test they share, and B_0 will also be produced. Here B_0 is the unweighted mean of means.

[1]Of course, all lines that are not parallel will cross at some point. In a disordinal crossed interaction they cross within the observed data.

[2]Note that these contrast codes satisfy the three criteria described in Section 8.5. Also note that these codes are perfectly correlated with unweighted effects codes for the two-group case—a desirability that will be discussed in the next section.

Contrast codes are particularly useful for interactions involving nominal scales because they are designed to be orthogonal and to represent meaningful differences between means of particular groups or combinations of groups. They allow the analyst to test the focal questions of interest whether there are two (or more) levels of each factor. These characteristics greatly facilitate the interpretation of interactions. However, other coding schemes may also be used, provided that the investigator keeps firmly in mind the meaning of both zero and a 1-unit change in each B_i in such sets.

To illustrate, suppose that the two research factors in Table 9.1.1 had been dummy-variable coded, with 0 assigned to the placebo and control conditions. In such a case X_3 the interaction, created as the product of the two main effects, would be coded 1 for the drug-frontal group ($X_1 = 1$ times $X_2 = 1$) and 0 for the other three groups. This results in the following set of code variables:

Treatment condition	X_D	X_S	X_{SxD}
Active drug, frontal	1	1	1
Active drug, sham	1	0	0
Placebo, frontal	0	1	0
Placebo, sham	0	0	0

As was shown for the interaction terms involving uncentered quantitative variables in Chapter 7, the results of this analysis may be interpreted. The significance test of the interaction term (X_{SxD}) and its partial correlation is not affected by the use of this dummy coding scheme. However, the absence of centering makes the average (first-order) effects awkward to interpret. The three data sets from Table 9.1.1 are shown analyzed using this dummy variable coding scheme in Table 9.1.2.

We can recreate the cell means by noting their representation in the three IVs and the intercept. The active drug-frontal group is coded 1 on all three variables—the dummy variable representing active drug, the dummy variable representing frontal lesion, and the product of these two variables—and we will need to add in the intercept in order to recreate its mean. Therefore $M_{DF} = B_1 + B_2 + B_3 + B_0$. The active drug-sham group is coded 1 on the drug dummy variable but 0 on the other IVs, so that its mean will $= B_2 + B_0$. The placebo-frontal group is coded 1 on the lesion dummy variable but 0 on the other IVs so that its mean

TABLE 9.1.2
Examples With No Interaction, Crossed Interaction, and Ordinal Interaction

	A. No interaction	B. Crossed interaction	C. Ordinal interaction
Dummy-variable analyses omitting the interaction:			
Frontal-sham difference $= B_1$	$8-6=2$	$12-8=4$	$10-6=4$
Active-placebo difference $= B_2$	$5-9=-4$	$9-11=-2$	$5-11=-6$
B_0	8	10	9
Analyses including the dummy variable product:			
$M_{PF} - M_{PS} = B_1$	$10-8=2$	$10-12=-2$	$12-10=2$
$M_{AS} - M_{PS} = B_2$	$4-8=-4$	$4-12=-8$	$2-10=-8$
$M_{AF} - M_{AS} - M_{PF} + M_{PS} = B_3$	$6-4-10+8=0$	$14-4-10+12=12$	$8-2-12+10=4$
$B_0 = M_{PS}$	8	12	10

will $= B_1 + B_0$. And finally, the placebo-sham group is coded 0 on all IVs so that its mean will $= B_0$. Thus, the regression coefficients will have the interpretation given in Table 9.1.2B.[3] From this exercise we see that using dummy codes and taking the product to represent the interaction leads to regression estimates that, on the whole, are unlikely to represent the actual hypotheses for which the analyses were carried out. The proper interaction effect is preserved, but all average effects and the intercept differ from the researcher's hypotheses. For example, B_1 is now the difference between the means of the placebo-frontal and the placebo-sham (reference) groups rather than the average difference between the means of the active drug and placebo conditions. This example illustrates the importance of choosing the appropriate coding system to represent one's research hypotheses.

The Relationship of Coding Method to CI and Significance Tests for Nominal Scale Interactions

To carry our fictitious example yet further, we test the significance of the estimated coefficients in models using the two alternative coding schemes in Table 9.1.3. The first table of each pair tests the average effects and interaction using contrast codes, and the second tests the same data using dummy-variable codes.[4] As we have noted, the variables in the two models do not, in general, test the same effects. Nevertheless, two comparisons are particularly relevant to a full understanding of the differences. The first is the overall R for the equation estimating Y from the three predictors, which is precisely the same in the two representations of these nominal variables, as we saw in Chapter 8 for the general case of coding g groups with $g - 1$ variables.

The second thing to note in these comparisons of dummy-variable and contrast coding of nominal scales and their interaction is that the significance test (and thus, necessarily, the statistical significance of sr and pr, not shown) for the interaction term has precisely the same t value, and therefore our confidence in its departure from zero in the population is precisely equal in the two cases. It happens in this example that B_3 also takes on the same value, although this equivalence will not hold in general. This is, of course, another special case of the principle presented in Section 7.2.6, that the significance of interactions is invariant over linear transformations of the variables. When not provided by output from computer programs, sr may be determined by

$$sr_i = \frac{\beta_i}{\sqrt{1/\text{tolerance}}}. \tag{9.1.1}$$

A third thing to note in the comparison of the dummy variables and contrast-coded variables is the change in the tolerance of the variables depending on the coding scheme. The tolerance ($= 1 - R^2_{i.12\ldots(i)\ldots k}$) is the proportion of an IV's variance that is independent of the other IVs and thus a measure of collinearity, which we will consider in more detail in Section 10.5.3. When the tolerance is equal to 1.0 the variable is uncorrelated with the other IVs. As the tolerance decreases, there is increasing overlap between X_i and the other IVs, and thus increasing difficulty in interpreting the meaning of the coefficients. In this case, as usual, the dummy-variable codes are substantially intercorrelated, as shown by the tolerance. The contrast codes, however, are nearly independent, all nonessential collinearity having been eliminated. Only essential collinearity that is due to the unequal numbers of cases in the different cells of our example keeps this value from being equal to 1.0.

---

[3]The reader may carry out the calculations to see that these equations do in fact recreate the means given in Table 9.1.1.

[4]For these tables we have not rounded the estimates to integers, as we had done in the earlier tables.

TABLE 9.1.3
Statistical Tests of 2 by 2 Data With and Without Interactions by Two Coding Methods

Effect	Value	SE_B	CI_B	Tolerance	t	p
(A) No interaction data: Contrast codes, $R = .913$						
B_0	7.01	.08	6.85–7.17	—	86.08	<.01
B_1	2.01	0.16	1.69–2.33	.99	12.34	<.01
B_2	−4.00	0.16	−3.68–(−4.32)	.95	−24.55	<.01
B_3	.03	0.33	−.30–(+.34)	.96	0.10	.92
No interaction data: Dummy-variable codes, $R = .913$						
B_0	8.01	0.138	7.74–8.28	—	58.05	<.01
B_1	1.99	0.23	1.54–2.45	.50	8.65	<.01
B_2	−4.01	0.21	−3.61–(−4.42)	.60	−19.52	<.01
B_3	.03	0.33	−.61–(+.68)	.36	0.10	.92
(B) Crossed interaction data: Contrast codes, $R = .970$						
B_0	10.01	0.08	9.85–10.17	—	122.94	<.01
B_1	4.01	0.16	3.69–4.33	.99	24.62	<.01
B_2	−2.00	0.16	−2.32–(−1.68)	.95	−12.27	<.01
B_3	12.03	−.33	5.70–6.34	.96	36.96	<.01
Crossed interaction data: Dummy-variable codes, $R = .970$						
B_0	12.01	0.14	11.74–12.28	—	87.03	<.01
B_1	−2.01	0.23	1.55–2.46	.50	−8.72	<.01
B_2	−8.01	0.21	−8.42–(−7.61)	.60	−38.98	<.01
B_3	12.03	0.33	11.39–12.68	.36	36.96	<.01
(C) Ordinal interaction data: Contrast codes, $R = .969$						
B_0	8.00	0.08	7.84–8.16	—	99.85	<.01
B_1	4.02	0.16	3.70–4.34	.99	25.07	<.01
B_2	−5.99	0.16	−6.39–(−5.67)	.95	−37.35	<.01
B_3	4.01	0.32	1.69–2.32	.96	12.52	<.01
Ordinal interaction data: Dummy-variable codes, $R = .969$						
B_0	9.99	0.14	9.72–10.26	—	73.53	<.01
B_1	2.01	0.23	1.56–2.46	.50	8.87	<.01
B_2	−7.99	0.20	−8.39–(−7.59)	.60	−39.49	<.01
B_3	4.01	0.32	3.38–4.65	.36	12.52	<.01

Lessons Learned from the 2 by 2 Example

From this simple example we may draw several conclusions that may be applied to more complex interactions involving nominal scales. They are:

1. Analysts should choose a coding scheme that most adequately represents their research questions. In experimental designs, this will typically be unweighted effects or contrast coding.
2. The significance tests on the total interaction effect for nominal scales,[5] as for continuous scales, will be invariant. In exploratory analyses, a hierarchical entering of IV sets into the regression equation can be used to test the contribution of the interaction set. This is likely to be an especially attractive option if one is merely checking the interactions to make sure that

[5]Taken collectively, as we shall see later. In this example there was only one interaction term.

they are not needed in the model, as in the ANCOVA model presented in Section 8.7. That is, if we have not expected or theorized an interaction, we may code the main effects in any way that makes most sense for the investigative purposes. We then use the products of these main effects collectively to test the significance of the interaction, expecting that the interaction terms will not add significantly to prediction and therefore can be omitted from the final equations to be reported. Note that the regression coefficients from the initial steps of the hierarchical analysis will differ from those in the final model. If the interaction term turns out to be significant, then the regression coefficients from the full model including the interaction should be reported. The analyst should also revisit whether the coding scheme continues to represent the research questions of interest in light of the changes in the interpretation of the lower order regression coefficients in the presence of an interaction.

3. Regardless of the coding scheme for nominal variable main effects and interactions, providing that they include $g - 1$ IVs, where g equals the number of groups (cells) in the design, and none of these $g - 1$ IVs is perfectly predictable from the others, the multiple R and R^2 will be invariant. In a two-factor design, g will equal the product of the number of levels of the first factor times the number of levels of the second factor. Such an invariance, in combination with the invariance of the interaction contribution to R^2, means that the contribution of the combined main effects is also invariant over alternative methods of coding.

4. The values of the regression coefficients for lower order terms that are part of higher order interactions may change dramatically when the coding scheme changes, and must be interpreted as the effect of each factor only at zero on the other factor(s).

9.1.2 Regression Analyses of Multiple Sets of Nominal Variables With More Than Two Categories

There are two distinct aspects of the analysis of k groups by k groups designs. The first is the determination of which research factors, including interactions, should be included in the full model. This decision is often made on the basis of the significant contribution of sets of variables to the overall prediction, as discussed in Chapter 5. Thus this analysis may determine whether certain sets of variables or interactions may be dropped from further consideration.

With equal ns in each condition of the design, the various main effect and interaction sets of independent variables are uncorrelated or orthogonal. When cell ns are proportional these sets are also orthogonal. By proportional, we mean that the number of cases in any cell can be exactly determined by multiplying the proportion of cases in that cell for each research factor by the proportion in that cell for every other research factor. Thus, if half the cases were in category 1 for research factor B and one fourth of the cases are in category 2 for research factor C, then the proportion of the full sample that is in cell B_1C_2 should be $.50 \times .25 = .125$. In this case a χ^2 on the table of sample ns will equal 0. Such a design is said to be balanced.

When the study design is fully balanced, with equal or proportional ns in all cells so that research factors are uncorrelated, the analysis may proceed by hierarchically examining sets representing research factors and interactions coded by whatever system best reflects the major research hypotheses. When the cell ns are unequal, however, it will matter how these codes are assigned to groups: Significance tests will vary as a function of these codes.

Often this first step will be considered essential before proceeding with the second task that is selecting the optimal code system for the categorical variables to represent the major research questions in the analytic output.

Significance Tests of Research Factors and Interaction Sets: Type I, II, and III Regression Sums of Squares

We will present the three approaches to testing the significance of group differences by considering the situation in which research factor A (with, e.g., four groups) is represented by a set of three variables $\boldsymbol{A}$, research factor B (with, e.g., three groups) is represented by a set of two variables $\boldsymbol{B}$, and the interaction between these sets is represented in this case by six variables, set $\boldsymbol{A} \times \boldsymbol{B}$. Type I sum of squares uses a hierarchical build up approach to estimate the overall significance of each effect, as presented in Section 5.4. Assume that A is the effect of most interest. Then three hierarchical regression equations are estimated, using $\boldsymbol{A}$ in the first, $\boldsymbol{A}$ and $\boldsymbol{B}$ in the second, and $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{A} \times \boldsymbol{B}$ in the third. The test of $\boldsymbol{A}$ is taken from the first equation, the test of $\boldsymbol{B}$ uses the gain in R^2 (increase in regression SS) from the second equation, and the test of $\boldsymbol{A} \times \boldsymbol{B}$ uses the gain in prediction from the third over the second equation. Recall that the gain in prediction in a hierarchical regression model is tested using either Model 1 error [Eq. (5.5.1)] in which each effect is tested for significance when it is first entered, or Model 2 error, in which the MS for the error is taken from the final equation. With this method the test of $\boldsymbol{A}$ ignores any possible overlap of $\boldsymbol{A}$ with $\boldsymbol{B}$ or the interaction, and the test of $\boldsymbol{B}$ controls for the effect of $\boldsymbol{A}$, but ignores the effect of the interaction. The decision about the sequence of $\boldsymbol{A}$ and $\boldsymbol{B}$ is made on the substantive grounds of most interest or presumed causal priority. The methods of coding $\boldsymbol{A}$ and $\boldsymbol{B}$ do not affect these tests. Even if the interaction were the effect of most interest, it would be tested at step 3: Interactions are partialed effects (see Chapter 7; J. Cohen, 1978).

Type II sum of squares uses a modified hierarchical approach. In addition to the hierarchical sequence used in Type I sum of squares we also estimate Y from an equation using only $\boldsymbol{B}$. The effect of $\boldsymbol{B}$ is determined as before, by the gain in prediction from the equation using only $\boldsymbol{A}$ to the equation using both $\boldsymbol{A}$ and $\boldsymbol{B}$. The effect of $\boldsymbol{A}$ is determined in a parallel manner, by the gain in prediction from the equation using only $\boldsymbol{B}$ to the equation using both $\boldsymbol{A}$ and $\boldsymbol{B}$. The test of the interaction is as before, its contribution to R^2 above the main effects. Again, the Model 1 or Model 2 error term may be employed, usually depending on whether there is an a priori reason for expecting an interaction effect.

The Type III sum of squares approach compares the prediction of the full model to submodels in which only the effect of interest is eliminated. This model is that used by current ANOVA programs applied to unequal n designs, and it depends critically on using unweighted effects codes (Section 8.3) to represent research factors $\boldsymbol{A}$ and $\boldsymbol{B}$ and their products to represent their interaction. One begins by estimating the "full model" equation for Y from $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{A} \times \boldsymbol{B}$. Then an equation employing $\boldsymbol{B}$ and $\boldsymbol{A} \times \boldsymbol{B}$ is used to determine the difference in the regression SS between the full model and the model omitting $\boldsymbol{A}$, and the resulting MS (dividing that difference by the *df* for $\boldsymbol{A}$) is tested using the full equation residual MS, as in Eq. (5.5.2). A parallel set of procedures, starting with an equation employing $\boldsymbol{A}$ and $\boldsymbol{A} \times \boldsymbol{B}$ is used to test the independent contribution of $\boldsymbol{B}$. Finally, the $\boldsymbol{A} \times \boldsymbol{B}$ effect is tested by determining the difference in regression SS between the combined $\boldsymbol{A}$ and $\boldsymbol{B}$ effects and the full model equation. This test is necessarily equivalent to the tests using either Type I or Type II SS approaches. The Type III sum of squares approach provides a test of the unique effect of each research factor with any contribution due to unequal cell ns partialed out. For this reason, it is viewed as the most conservative approach. It is the same approach used when we test the significance of each predictor in the full equation, providing that we have used unweighted effects codes.

As we saw in Chapter 5, researchers should choose the approach that best represents their central questions of interest. A critical prerequisite in making this choice is to consider whether a random or representative sample has been selected from a population. The Type I sum of squares approach implies the strong assumption that the differences in sample sizes represent differences in the proportion of cases in each of the conditions in the population. Thus, it is not appropriate for experimental studies in which investigators (and perhaps chance factors) determine cell sizes that the Type I SS be used. Differences in sample sizes across conditions almost always represent either procedural decisions by the experimenter (e.g., using fewer participants in a particularly difficult to implement or costly treatment condition), a failure of the randomization procedure to achieve equal allocation of the participants to treatment conditions, or missing data in one or more of the treatment conditions (see Chapter 11 for a discussion of missing data issues). For this reason, sources emphasizing the analysis of experiments discourage the use of the Type I sum of squares approach (e.g., Kirk, 1995).

Given a random or representative sample, there are a number of research contexts in which the Type I sum of squares approach provides the optimal approach. For example, imagine a sociologist is studying the effects of family socioeconomic status (low versus middle versus upper) and high school graduation (no versus yes) on lifetime income. The researcher may theorize that family socioeconomic status (SES) is a cause of successful high school graduation so that any overlapping variance between these two IVs should be assigned to SES. Howell and McConaughty (1982) offer several illustrations of contexts in which both theoretical and applied policy questions may be optimally answered using the Type I sum of squares approach.

The choice between the Type II and Type III sum of squares approaches is normally based on the researcher's assumption about the existence of the interaction term in the population. If the effect of the interaction set is assumed to be 0, then the Type II sum of squares approach provides a more powerful test of the main effects of A and B. However, if the $A \times B$ interaction is not 0 in the population, then the model is not correctly specified and the estimates of A and B will be biased. We previously considered this general issue of bias versus efficiency in model specification in Chapter 4. A reasonable approach in the absence of an expected interaction effect is to test the significance of the interaction effect and, if it is not significant, to delete it from the model. The usefulness of this approach will also depend on the power of the test of interaction—the interaction may exist in the population, but may not be detected in a small sample.

Designs involving more than two factors will involve straightforward extensions of these procedures. Maxwell and Delaney (1990, Chapters 7 and 8) present a full and balanced discussion of issues in the analysis of factorial designs with unequal sample sizes.

Selection of Group Coding Method and Significance Tests of Individual Variable Effects

Having determined the significance of the overall effect of each research factor and interactions among research factors, the next task is determination of the best coding system to represent the substantive issues in the full equation model. In a sense this is the payoff of analyzing the data using an MRC rather than an ANOVA approach, in addition to the flexibility of including other, noncategorical, variables as appropriate to the substantive issues. For these decisions we refer back to the options presented in Chapter 8. We also note that one may employ an unweighted effects code system in order to assess the unique contribution of a set, using Model III SS as above and yet switch to some other method of coding to provide the best possible answers to the substantive questions motivating the research.

Illustrative Example

Consider a study comparing the efficacy (Y) of three different treatment procedures (T), two experimental and one control. The researchers succeed in getting each of the four medical school hospitals in a large metropolitan area to participate. The three different procedures are CH09EX02
employed on randomly assigned samples of suitably selected patients at each of the four medical school hospitals (H: H_1, H_2, H_3, H_4), thus making possible an appraisal of the uniformity of the treatment effects across the medical school hospitals. Since the four hospitals constitute the

TABLE 9.1.4
Effects of Research Factors and Interaction in an Experimental Study

	SS	*df*	MS	*F*
IV sets				
$H, T, H \times T$	75,209.071	11		
Residual$_{H,T,H\times T}$	30,891.422	96	321.79	
$H, H \times T$	52,193.908			
$T, H \times T$	71,077.543			
H, T	30,820.806			
Unique contribution				
H	75,209.071	3	1377.18	4.28
	−71,077.543			
	4,131.528			
T	75,209.071	2	11,007.58	34.21
	−53,193.908			
	22,015.163			
$H \times T$	75,209.071	6	7,398.04	22.99
	−30,820.806			
	44,388.265			

entire population of hospitals of interest, this is a 4 by 3 factorial design, with the interaction (***H*** by ***T***) directly addressing the issue of uniformity of effects.[6] Each of the 12 cells contains efficacy scores for the patients treated by one of the three procedures at one of the four hospitals, and in the interest of generality, there are a different number of cases in each of the 12 subgroups. The n for the entire study is 108.

Given the unequal *n*s in this example, and the fact that it includes an experimental component, our first task will be to test the overall contribution of between hospital differences, treatments, and their interaction by means of the Type III SS method. Thus we will code the three variables representing the four hospitals H and the two variables representing the three treatment conditions T by unweighted effects codes. For this purpose it is immaterial which group is selected as the contrast group, the estimated SS will remain the same. The regression SS for the full equation including $\boldsymbol{H}, \boldsymbol{T}$ and $\boldsymbol{H} \times \boldsymbol{T}$ is 75209.071, and the residual MS from that equation = 321.79. To test each of the individual factor sets $\boldsymbol{H}, \boldsymbol{T}$ and $\boldsymbol{H} \times \boldsymbol{T}$ we take the difference between the full equation regression SS and the regression SS for all other sets, as shown in Table 9.1.4. In this example all three sets are statistically significant ($p < .01$).

Having determined the importance of all three sets, we now turn to the task of optimal coding of the individual variables to represent our major research questions in the full equation. Presuming that we have no special interest in any single one of the hospitals our coding method should treat them all on an equal footing, so that any $\boldsymbol{H}$ involves comparisons among the four hospitals all on equal footing, and any comparison conveniently proceeds as being between the Y mean of any given hospital and the unweighted mean of means of all four hospitals. Because there are four hospitals and only three *df*, one hospital must be selected as the base group and coded consistently −1. Table 9.1.5 provides the effects codes for the four hospitals as X_1 to X_3.

[6]Technically, hospitals is a fixed effect in the present design because our conclusions are intended to apply only to these three treatments and these four hospitals. In cases in which many hospitals are randomly selected from a population of hospitals, multilevel modeling techniques described in Chapter 14 should be employed.

For the T factor, on the other hand, the presence of a control group to which each of the experimental groups is to be compared suggests the possible use of dummy-variable coding (Section 8.2). The three treatments are thus represented by two IVs (X_4 and X_5) with the control condition consistently assigned 0, as shown in Table 9.1.5. Thus, each patient is characterized on hospital of origin by three IVs and for treatment group by two IVs, which together produce the coding of X_1 through X_5 in Table 9.1.4.

The interaction set $\boldsymbol{H} \times \boldsymbol{T}$ is again created by multiplying each of the $\boldsymbol{H}$ set variables by each of the $\boldsymbol{T}$ set variables, creating $2 \times 3 = 6$ new IVs. The 11 IVs in the three sets exactly identify the H_iT_j cell of each patient. We emphasized in Chapter 8 that to fully represent G, made up of g groups, $g - 1$ IVs are necessary, whatever the form of coding. This 4 by 3 design results in 12 groups, hence full representation requires 11 IVs. From the many optional methods of coding membership in one of 12 groups, the coding given in Table 9.1.5 represents these 12 groups as three effects-coded IVs for $\boldsymbol{H}$, two dummy-coded IVs for $\boldsymbol{T}$, and the six interaction-bearing IVs that result from their set by set multiplication. As we will see, designs in which two different coding systems are employed require special care to assure proper interpretation of the individual regression coefficients.

TABLE 9.1.5
Codes for the *H* by *T* (4 by 3) Factorial Design

(A) $\boldsymbol{H}$ (unweighted-effects codes)

	X_1	X_2	X_3
H_1	1	0	0
H_2	0	1	0
H_3	0	0	1
H_4	−1	−1	1

(B) $\boldsymbol{T}$ coding (dummy codes)

	X_4	X_5
T_1	1	0
T_2	0	1
Control	0	0

(C) Joint coding of the 12 cells of the $\boldsymbol{H}$ by $\boldsymbol{T}$ factorial design

	Hospitals			Treatments		Hospitals by treatment					
						X_1X_4	X_1X_5	X_2X_4	X_2X_5	X_3X_4	X_3X_5
Cell	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
H_1T_1	1	0	0	1	0	1	0	0	0	0	0
H_1T_2	1	0	0	0	1	0	1	0	0	0	0
H_1C	1	0	0	0	0	0	0	0	0	0	0
H_2T_1	0	1	0	1	0	0	0	1	0	0	0
H_2T_2	0	1	0	0	1	0	0	0	1	0	0
H_2C	0	1	0	0	0	0	0	0	0	0	0
H_3T_1	0	0	1	1	0	0	0	0	0	1	0
H_3T_2	0	0	1	0	1	0	0	0	0	0	1
H_3C	0	0	1	0	0	0	0	0	0	0	0
H_4T_1	−1	−1	−1	1	0	−1	0	−1	0	−1	0
H_4T_2	−1	−1	−1	0	1	0	−1	0	−1	0	−1
H_4C	−1	−1	−1	0	0	0	0	0	0	0	0

TABLE 9.1.6
M_Y and n_i in the 4 by 3 Nominal Scale Example

	T_1	T_2	T_3 = control	M_H	M_{mT}
H_1	47.8 (9)	123.9 (9)	86.6 (6)	86.0 (24)	86.1
H_2	71.2 (10)	134.5 (8)	59.5 (9)	86.1 (27)	88.4
H_3	54.7 (12)	95.5 (9)	97.9 (9)	79.9 (30)	82.7
H_4	93.9 (9)	53.9 (6)	68.0 (12)	73.5 (27)	71.7
M_T	66.1 (40)	105.4 (32)	76.4 (36)	81.2 (108)	
M_{mH}	66.7	102.0	78.0		82.2

Given the efficacy scores (Y) for the 108 patients, together with the representation of each patient as to hospital H and treatment group T membership using the 11 IVs of Table 9.1.5, the data matrix is completely defined.

The full equation with IV sets $\boldsymbol{H}$, $\boldsymbol{T}$ and $\boldsymbol{H} \times \boldsymbol{T}$ is

$$\begin{array}{lllllllll}
\hat{Y} = 8.60X_1 & -19.55X_2 & -19.94X_3 & -11.29X_4 & +23.95X_5 & -27.49X_6 & +13.38X_7 & +22.99X_8 \\
t\ 1.43 & -3.54 & 3.81 & -2.69 & 5.38 & -3.48 & 1.67 & 3.20 \\
+51.05X_9 & -31.92X_{10} & -26.37X_{11} & +78.01 \\
t\ 6.72 & -4.56 & -3.54
\end{array}$$

and the t tests for these effects are provided beneath each term. With 96 error df a $t \geq 2.63$ is significant with $\alpha = .05$. Each of these unstandardized coefficients represents a specific effect. B_1 represents the difference between the mean of hospital 1 and the unweighted mean of the four hospitals in the control condition. Table 9.1.6 provides the full means and ns for each cell in these data where it can be seen that this coefficient, 8.60, does in fact = 86.6 − 78.0. B_2 and B_3 represent similar comparisons for hospitals 2 and 3, respectively. B_4 represents the comparison of the mean for T_1 with the mean of the control group. The mean for T_1 is the unweighted mean of the T_1 groups across the four hospitals and the mean for the control is the unweighted mean of the control groups across the four hospitals. B_5 represents a similar comparison of the mean of T_2 with the mean of the control group. Each of the six interaction terms represents a specific facet of the nonuniformity of effects. Each of these IVs represents a specific effect aspect of $\boldsymbol{H}$ by dummy aspect of $\boldsymbol{T}$, its B gives the size of this discrete interaction effect, and the accompanying t provides its significance test. Five of these six are significant in this fictitious example. The first is X_6, the interaction between the effect of hospital 1 (relative to the unweighted mean of the hospitals) and the treatment 1 comparison with the control. The T_1 versus control comparison in H_1 is 47.8 − 86.6 = −38.8. The T_1 versus control at the unweighted mean of hospitals is 66.7 − 78.0 = −11.3. The difference between these (the interaction) is −38.8 − (−11.3) = −27.5, the B for the first interaction term, X_6. The second interaction term tests the (nonsignificant) difference between T_2 and control in H_1 (again as compared to the mean of the hospital mean differences), and the third term X_8 compares the H_2T_1 versus control mean difference (71.2 − 59.5 = 11.7) with the same T_1 versus control difference at the unweighted mean of the hospitals = 11.7 − (−11.3) = 23. The remaining three terms can be similarly interpreted.

9.2 INTERACTIONS INVOLVING MORE THAN TWO NOMINAL SCALES

In Chapter 7 we saw that when the interaction among more than two continuous variables is being examined it is necessary to include in the equation all of the lower order interactions for the model to be hierarchically well specified (see Section 7.8; Peixoto, 1987). Otherwise, the interpretation of the terms in the regression equation will be confounded by any absence of lower order terms. Similarly, when nominal scales are each represented as a set of $k - 1$ code variables, where k is the number of categories (groups) that comprise the nominal variable, interactions involving more than two such nominal variables will require inclusion of the full set of code variables representing each possible lower order interaction and the original nominal variables. Thus, examination of the triple interaction between three categories of ethnicity E, three religious affiliations R, and three school grades S (e.g., 6th, 8th, and 10th) in predicting Y would require, in addition to $E_1, E_2, R_1, R_2, S_1, S_2$, and the triple interaction set of eight variables $E_1 \times R_1 \times S_1, E_1 \times R_2 \times S_1, E_1 \times R_1 \times S_2, E_1 \times R_2 \times S_2, E_2 \times R_1 \times S_1, E_2 \times R_2 \times S_1, E_2 \times R_1 \times S_2, E_2 \times R_2 \times S_2$, the three sets of two-way interactions collectively having another 12 variables, for a total of $6 + 12 + 8 = 26$ variables representing the 27 cells of this 3 by 3 by 3 design.[7] Necessarily such an investigation would require either a very large sample or very large expected effect sizes to detect such differences with reasonable power and confidence limits on estimates. Thus it is most frequent that researchers will investigate such interactions only when there is both a strong theoretical reason for interest and a large sample to assure adequate statistical power. The fact that several nominal variables exist in a data set and can be used to create a large number of different potential two-way, three-way, four-way, or higher way interactions does not necessarily mean that they are all to be investigated in the absence of compelling research questions.

As we have noted, when cell sizes are unequal the different group factors will be correlated. Thus, the first step will be to use an appropriate method of testing the contribution of group and interaction sets. In experimental studies this will most readily be accomplished by employing a standard ANOVA program, which will use unweighted effects codes to estimate these effects. Alternatively, effects codes can be created by the investigator and the Type III Regression SS method described earlier carried out. In other research designs the investigator may choose a Type I Regression SS method, using hierarchical sequences of main effect and interaction sets, or a Type II Regression SS method, using multiple hierarchical sequences of main effect and interaction sets.

Once the determination of the significance of sets and the level of interaction to be retained in the final model is made, the investigator proceeds to the decision regarding coding methods to employ in the full equation model.

9.2.1 An Example of Three Nominal Scales Coded by Alternative Methods

We present here a fictitious data set based in part on the theorizing and experimental findings of Carol Dweck (1999). The dependent variable of interest is the amount of effort the individual exerts in solving a set of problems in the *second* part of an experimental session. The manipulations take place in association with the participant's work on an experimental task given during the first part of the experimental session. The first factor is task difficulty (D)—the experimental task is either hard or easy. The second factor is feedback (F)—participants are told that they have either succeeded or failed on the experimental task. The final factor is attribution (A) with three levels. Participants are led to believe that the outcome of their performance was due to their effort (A_1) their ability (A_3), or they are given no basis for making an attribution about the cause of their performance (control group, A_2). Thus, we have a 2 (D) by 2 (F) by 3 (A) design in which each of the $2 \times 2 \times 3 = 12$ possible treatment conditions are represented.

---

[7]This exposition assumes no empty or nearly empty cells in the design. If there are empty cells some appropriate simplification is required. Typically, researchers consider collapsing categories within a nominal category that they do not expect to differ or dropping consideration of the three-way and possibly some of the two-way interactions from the model.

TABLE 9.2.1
Hierarchical Analysis of Effects-Coded Variables in a Three Factor Example

Predictor sets	SS	df	MS	R^2	F	p
Main effects	612.17	4	153.04	.350	15.47	<.01
Main effects + two-way interactions	890.60	9		.509		
Full model	909.50	11	82.68	.520	10.62	<.01
Residual	840.41	108	7.78			

A central tenet of Dweck's theorizing is that important consequences follow from viewing task successes and failures as resulting from a relatively fixed ability ("trait" attribution) or a function of the degree and kind of effort put into the task ("effort" attribution). The theory states that the effect of this *attribution* variable on effort (Y) will depend on the previous success or failure experience with a comparable task, with the "trait" group more likely to give up and put less effort into a task that is similar to a previously failed task, whereas the "effort" group is more likely to put more effort into such a task. Furthermore, this interaction may also depend on the difficulty of the task. Our fictitious example has unequal *n*s in the various cells, which makes it as general as possible.

We first determine the significance of main effects and interaction sets in the full model. Given the unequal *n*s and the experimental nature of this study, for this analysis we may either use an ANOVA computer program or, equivalently, code all sets by unweighted effects codes and their products and carry out a Type III Regression SS analysis.

Carrying out these analyses by a "step-down" procedure, starting from the full model CH09EX03 including attribution $= \boldsymbol{A}$ with $k = 2$, difficulty $= \boldsymbol{D}$ with $k = 1$, and failure $= \boldsymbol{F}$ with $k = 1$; $\boldsymbol{A} \times \boldsymbol{D}, \boldsymbol{A} \times \boldsymbol{F}, \boldsymbol{D} \times \boldsymbol{F}$ two-way interactions with $k = 5$, and 2 three-way interaction variables, $A_1 \times \boldsymbol{D} \times \boldsymbol{F}$ and $A_2 \times \boldsymbol{D} \times \boldsymbol{F}$. Table 9.2.1 provides these analyses.

In the first analysis comparing the prediction for all 11 variables with the prediction omitting the three-way interaction terms, we see that these two variables added only $909.50 - 890.60 = 18.90$ to the regression SS above the two-way interactions. This contribution of only about 1% to the prediction was not statistically significant. We decide to omit the triple interaction from further consideration, and proceed to investigate the two-way interaction contribution. For that model we see that the addition of the five two-way interaction terms added 278.45 to the regression SS over the main effects. Dividing by the 5 df = MS = 55.69, and testing this by the new residual MS $= 849.31/110 = 7.72$ yields $F = 7.21$, which with 5 and 110 df indicates $p < .01$. Of course the existence of two-way interactions indicates the necessity of including the main effects, and this model is accepted for the analysis of the study.

Variable Coding for Analysis of the Full Model

A second consequence of the potentially large number of total variables when more than two-way interactions among sets representing nominal scales are investigated is the critical importance of the selection of the method of coding the variables. As we saw in the previous example, when interactions among nominal scales are being examined, as elsewhere, it is important that the meaning of the specific statistical tests being made is pertinent to the purposes of the investigation; otherwise the individual t tests may refer to comparisons that are not of any interest. Such a consideration is at least equally important when one is considering interactions among more than two nominal scales or, as we shall see subsequently, interactions between nominal and continuous scales. For pedagogical purposes, we will use each of the four major coding systems to code all three factors in the analyses presented here. However, as we saw in the earlier example investigating treatments at different hospitals, the investigator's hypotheses should determine which coding scheme is used for each of the nominal IVs. Such hypotheses may indicate different coding approaches to the different nominal scales in a single analysis.

Table 9.2.2 presents the details of the four coding schemes. The particular coding system that is employed is indicated by the addition of (D), (U), (W), or (C) to the subscript for each of the code variables. For example, $X_{1(D)}$ represents the first dummy code which involves the comparison of the effort attribution group with the control group. As an overview, the first system is dummy-variable coding, where we must select one of the groups for each nominal scale as the reference group against which the other groups will be compared. In this case we have selected A_2, the control group for attribution, as well as the difficult task condition (D_2), and the failure outcome condition (F_2). The second set of codes represent the unweighted effects codes, where we have chosen for the base group for each factor the same groups that were used as the reference groups for the dummy codes. Following the practice in unweighted effects codes, each base group is coded -1. The third coding scheme is unweighted effects coding which is included here only for illustrative purposes. As noted in Section 8.4, this coding scheme is almost never appropriate for experimental work because it looks upon the actual cell sizes as being representative of their proportion in some population. It uses the number of cases in the group that makes up each nominal category to determine to the values of the code variables for the base group. Dweck's theory would predict a specific pattern of means as a function of the attribution condition in the amount of effort the individual exerts in solving the problems in the second part of the experiment ($M_{A1} > M_{A2} > M_{A3}$). The difference in the means of the attribution groups would be expected to be greater in the failure than in the success group, yielding an attribution $\times$ outcome interaction. The strong form of this hypothesis suggests that the $M_{A1} - M_{A2}$ difference is equal to the $M_{A3} - M_{A2}$ difference within each of the outcome conditions. This hypothesis is represented by code variable $X_{1(C)}$ that has values of $-0.5, 0$, and $+0.5$ for the ability, no attribution, and effort conditions, respectively. The second code variable $X_{2(C)}$ for A has values of $-0.5, 1$, and -0.5, and represents the nonlinear (quadratic) component of the differences among the groups. In the context of the strong version of the hypothesis, it is expected to be nonsignificant and serves as a check that the linear effect of attribution adequately represents the pattern of means in this data set. Given the $G_A - 1 = 2\,df$ for A these two contrasts fully represent the nominal variable. The contrast

TABLE 9.2.2
Alternative Codes for Main Effects in a Three-Nominal-Scale Example

Method	Dummy codes		Unweighted effects		Weighted effects		Contrast	
	$X_{1(D)}$	$X_{2(D)}$	$X_{1(U)}$	$X_{2(U)}$	$X_{1(W)}$	$X_{2(W)}$	$X_{1(C)}$	$X_{2(C)}$
Attribution set								
Group A_1	1	0	1	0	1	0	.5	−.5
Group A_2	0	0	−1	−1	$-n_{A1}/n_{A2}$	$-n_{A3}/n_{A2}$	0	1
Group A_3	0	1	0	1	0	1	−.5	−.5
Difficulty set	$X_{3(D)}$		$X_{3(U)}$		$X_{3(W)}$		$X_{3(C)}$	
Group D_1	1		1		1		.5	
Group D_2	0		−1		$-n_{D1}/n_{D2}$		−.5	
Failure set	$X_{4(D)}$		$X_{4(U)}$		$X_{4(W)}$		$X_{4(C)}$	
Group F_1	1		1		1		.5	
Group F_2	0		−1		$-n_{F1}/n_{F2}$		−.5	

codes for the three nominal variables all meet the three rules for contrast codes (see Section 8.5). Values on each code variable sum to zero, the code variables are linearly independent across variables [equivalently, the products of the coefficients across each possible pair of the four code variables ($X_{1(C)}, X_{2(C)}, X_{3(C)}, X_{4(C)}$) sum to zero], and the difference between the values of the codes for the two groups or combinations of groups on each variable equals 1.

For each of the coding schemes, there are three first-order (average) effects for the factors represented by the four code variables X_1, X_2, X_3, and X_4. Interactions between nominal variables are represented as products of the code variables representing the factors. There are three two-way interactions, which are represented by five code variables: the products of the two variables in A with the single code variables in D and F and the product of the D and F code variables. These five two-way interations are represented by five code variables X_5, X_6, X_7, X_8, X_9, and X_{10} in our example.

Table 9.2.3 provides the means and sample sizes for each of the cells in the design, as well as certain means of means. The coefficients produced by the alternative coding methods in this model are shown in Table 9.2.4. The first thing to notice is that the values, meaning, and significance tests for variables within sets are not equivalent across methods.

Taking these analyses in sequence, we note that the intercept B_0 for the full dummy-variable model must equal the mean of the group consistently coded 0 on all IVs, here the middle A group in the high difficulty and failure condition. In the model omitting the nonsignificant triple interactions, the intercept is only an approximation of that value. The meaning of other coefficients in the dummy-variable model is quite difficult to interpret because the high collinearity between the IVs thus coded requires consideration of most other coefficients in the interpretation of any one of them. In the full model tolerances for both main effects and interactions are less than 25%, and mostly less than 20%. Thus this method of coding is difficult to interpret, and the more so when sample sizes are unequal.[8]

The unweighted effects solution is the solution employed by ANOVA programs for unequal cell n, and is straightforward to interpret. B_0 in the full model equals the mean of all the cell means, and in the reduced model we have chosen is an approximation of this value.[9] The tolerances are much higher in this model, going little below 70%, indicating the greater simplicity of the model. As in ANOVA, the two-way interaction effects are tests of differences between differences. For example, the $X_{3(U)}$ by $X_{4(U)}$ interaction of .3 estimates the difference between the mean difference between easy and difficult conditions for the succeed condition $(11.9 - 11.2 = .7)$ and the mean difference between easy and difficult conditions for the fail condition $[10.9 - (-)11.0 = .1]$ be .3 times 2 (because the effects codes are 1 and -1) = .6.

TABLE 9.2.3
Cell Means and Numbers in the Three-Nominal-Scale Example

	Succeed = F_1		Fail = F_2			
	Easy = D_1	Difficult = D_2	Easy = D_1	Difficult = D_2	M_M	M_Y
Attribution$_1$ = Effort	13.04 (9)	12.57 (9)	14.80 (9)	16.70 (10)	14.28	14.34
Attribution$_2$ = None	11.23 (11)	10.43 (11)	12.28 (8)	10.14 (16)	11.01	10.84
Attribution$_3$ = Trait	11.49 (11)	9.60 (10)	6.54 (8)	6.07 (8)	8.43	8.74
M_M	11.92	10.87	11.21	10.97	11.24	11.27

[8] We have included on the disk a data file using approximately the same cell means but with equal cell sample sizes for the reader's examination. The dummy-variable solution is not really easier in this case, and tolerances are still less than .25.

[9] In the equal cell n condition, the estimated effects for main effects and interaction models do not change for the unweighted (or equivalent weighted) means solution because collinearity is only between IVs within sets.

TABLE 9.2.4
Table of B_i (SE), and t Values in Curtailed Model With Alternative Coding Methods

$B_i(SE)$, t	Dummy variable	Unweighted effects	Weighted effects	Contrast
B_0	10.5 (0.7)	11.2 (0.3)	11.2 (0.3)	11.2 (0.3)
B_{A1}	5.8 (1.0)	3.1 (0.4)	3.2 (0.4)	5.9 (0.7)
t	5.8**	8.3**	8.3**	9.0**
B_{A2}	−4.6 (1.1)	−2.8 (0.4)	−2.9 (0.4)	−0.3 (0.4)
t	4.3**	7.5**	7.4**	0.8
B_D	1.1 (1.0)	0.3 (0.3)	0.3 (0.3)	0.6 (0.5)
t	1.1	1.2	1.3	1.3
B_F	−0.6 (0.9)	0.2 (0.3)	0.2 (0.3)	0.4 (0.5)
t	0.6	0.7	0.6	0.7
$B_{A1 \times D}$	−2.2 (1.3)	−0.7 (0.4)	−0.7 (0.4)	−2.0 (1.3)
t	1.7	1.9	1.9	1.5
$B_{A2 \times D}$	−0.2 (1.3)	0.3 (0.4)	0.3 (0.4)	0.8 (0.7)
t	0.2*	0.8	0.7	1.1
$B_{A1 \times F}$	−2.8 (1.2)	−1.7 (0.4)	−1.8 (0.4)	−7.2 (1.3)
t	2.2**	4.5**	4.3**	5.5**
$B_{A2 \times F}$	4.4 (1.3)	1.9 (0.4)	2.1 (0.4)	−0.6 (0.7)
t	3.5**	5.2**	5.1**	0.8
$B_{D \times F}$	0.7 (1.0)	0.2 (0.3)	0.2 (0.3)	0.7 (1.0)
t	0.7	0.7	0.7	0.7

*The .05 α criteria has been met.
**The .01 α criteria has been met.

The weighted effects solution is intended to generalize to a population in which cell frequencies are not equal. The estimates are similar to the unweighted estimates to the extent that cell *n*s do not differ very substantially and, of course, are equivalent when they are equal. Because the cell *n*s within factors are not necessarily proportionate, it is generally not advisable to attempt to interpret a reduced model with this model, but the interpretation of the B_is is otherwise essentially comparable to those for the unweighted effects model, the difference lying in the population to which the generalization is desired.

The IVs in the contrast model are nearly orthogonal (exactly orthogonal in the equal cell n case), and thus both easier to interpret and more stable between the full and reduced models. We have designed them to emphasize the investigator's hypotheses. Once again, the intercept equals the mean of means. $X_{1(C)}$ gives us the mean difference in effort between the trait (A_1) and effort (A_3) groups across conditions (−5.8), $X_{2(C)}$ gives us the less interesting difference between the middle group (A_2) and the mean of the trait and effort groups. The interaction term B_5 contrasts the $A_1, D_1 + A_3, D_2$ means $= 13.9 + 7.8 = 21.7$ (averaged across F) with the $A_1, D_2 + A_3, D_1$ means $= 14.6 + 9.0 = 23.6$ (averaged across F) $= 1.9$ (in the full model, approximated by 2.0 in the model without triple interactions). Other effects can be similarly interpreted on the basis of the cell mean contrasts selected by the investigator.

The full set of cell means (or adjusted means if there are other variables in the model) can always be reproduced by summing across the B and code value products. For example, with dummy variables the means for the 12 cells estimated from the model without interactions are as shown in Table 9.2.5. The reproduced means omitting the triple interaction terms, as we have done here, are not exactly equal to the original means, although they clearly show the overall

TABLE 9.2.5
Recreating Adjusted Cell Means From the Regression Coefficients Using Dummy Variables

Cell	B_{A_1} 5.8 Code	B_{A_2} −4.6 Code	B_D 1.1 Code	B_F −0.6 Code	$B_{A_1 \times D}$ −2.2 Code	$B_{A_2 \times D}$ −0.2 Code	$B_{A_1 \times F}$ −2.8 Code	$B_{A_2 \times F}$ 4.4 Code	$B_{D \times F}$ 0.7 Code	B_0 10.5 Code	$\sum B_i \times$ Code
$A_1D_1F_1$	1	0	1	1	1	0	1	0	1	1	12.5
$A_1D_2F_1$	1	0	0	1	0	0	1	0	0	1	12.9
$A_1D_1F_2$	1	0	1	0	1	0	0	0	0	1	15.2
$A_1D_2F_2$	1	0	0	0	0	0	0	0	0	1	16.3
$A_2D_1F_1$	0	0	1	1	0	0	0	0	1	1	11.8
$A_2D_2F_1$	0	0	0	1	0	0	0	0	0	1	9.9
$A_2D_1F_2$	0	0	1	0	0	0	0	0	0	1	11.6
$A_2D_2F_2$	0	0	0	0	0	0	0	0	0	1	10.5
$A_3D_1F_1$	0	1	1	1	0	1	0	1	1	1	11.3
$A_3D_2F_1$	0	1	0	1	0	0	0	1	0	1	9.7
$A_3D_1F_2$	0	1	1	0	0	1	0	0	0	1	6.8
$A_3D_2F_2$	0	1	0	0	0	0	0	0	0	1	5.9

pattern. The reader may use the accompanying file to determine that if all 11 terms had been included the original means will be reproduced by summing the regression coefficient–code value products.

9.2.2 Interactions Among Nominal Scales in Which Not All Combinations Are Considered

Of course, if one is simply trying to reproduce the ANOVA with unequal cell numbers unweighted effects coding is mathematically equivalent to the method employed by most computer packages. And, as noted, one can reproduce the original cell means from any of the other coding systems. But there are much easier ways to reproduce the original cell means, and ANOVA and multiple comparison methods are tailored to test mean differences and differences between differences for statistical significance. When sets of variables representing nominal scales are used in MRC analyses, they are generally used in quite different ways than in ANOVA and ANCOVA. One purpose of nominal scale variable sets is serving as control variables in regression equations in which the effects of one or more quantitative variables are of major interest. Under these conditions and in many other situations the investigator may not be interested in or may not expect many of the potential interactions between the nominal and quantitative variables to contribute significantly to the prediction of Y. Although our focus was on the nominal variable, Section 8.7 considered these models. Yet another situation, to be reviewed in Section 9.3, is when there may be interactions between quantitative and nominal IVs so that the effects of some continuous variables may vary as a function of the categories that comprise the nominal IV.

There are many cases in which it is reasonable to examine certain interactions among nominal variables, but not all of them, either for reasons of parsimony, or in the effort to conserve degrees of freedom where hypotheses are not compelling. Most frequently such circumstances will arise in observational research, where potential interactions may be very many and clear a priori hypotheses may not exist. Occasionally, such circumstances may also arise in experimental studies. For example, an additional treatment factor or nominal individual difference variable (e.g., sex, ethnic group) may be included in the design to explore the generality of the hypothesized treatment effects.

To illustrate this issue, consider an observational study in which male and female (S) participants from three ethnic groups (E) are included in the study of a DV. The design is a 2 by 3 factorial design, and the investigator has clear hypotheses about that there will be overall sex differences and a S by E interaction. He anticipates that ethnic groups 1 and 2 will show clear sex differences in their mean scores on Y, but ethnic group 3 will show no sex differences. He uses two contrast codes for ethnic group with $E_1 = +1/3, +1/3$, and $-2/3$, and $E_2 = -0.5, +0.5$, and 0. In principle, the $S \times E_2$ interaction term may be omitted as not hypothesized or of interest. However, in general the first reasonable check will be to make certain that this variable does not contribute significantly to the prediction. The candidate term is tested and if it is *not* statistically significant, it is dropped from the model. It is prudent in such tests of possible model respecification to use a higher than usual value of the Type I error, say $\alpha = .20$, to help minimize the possibility that effects that in fact exist in the population (but which are not detected because of low statistical power) are inappropriately dropped from the regression equation. A complex regression equation must be "trimmed" from the highest order term down; one would not retain a three-way interaction term such as $\boldsymbol{A} \times \boldsymbol{B} \times \boldsymbol{C}$ but eliminate a two-way interaction term between included factors such as $\boldsymbol{A} \times \boldsymbol{B}$ because all lower order terms must be included in equations containing higher order terms. Aiken and West (1991, Chapter 6) discuss procedures for identifying the order in which terms may be trimmed in complex regression equations.[10]

In fact, many tests of interaction sets of all kinds are designed more as assumption checks or tests of the generality of one's findings of interest than as hypothesis tests. As considered in the previous example of treatments in medical school hospitals, it may be that the test of the hospital by treatment interaction was appropriately carried out only as a test of the assumption that such interactions were not needed. If the results had turned out that the interaction set as a whole did not contribute significantly to the prediction of Y, and the investigator had no a priori reason for thinking that it would, the terms representing the H by T interaction could be dropped from the regression model, although doing so would depart from the ANOVA model. Such tests of assumptions and of the generality of findings are typically briefly noted in research reports but are not fully presented.

If the investigator decides to leave out one or more of the highest order interactions between two or more nominal IVs, the choice of an appropriate coding system for the remaining variables will be quite critical. As we have noted throughout our presentation of nominal scale codes, it is not always intuitively obvious what the contrast will be when the regression model is modified. It follows, then, that it may also not be obvious what the consequences may be of omitting or adding one or more of a set of interaction variables. Because this will depend highly on the coding system, the most general advice is to beware of mistaken inferences drawn from incomplete sets. Each time the regression model is changed by adding or deleting interactions (or other higher order effects), the proper interpretation of each effect in the model needs to be revisited.

9.2.3 What If the Categories for One or More Nominal "Scales" Are Not Mutually Exclusive?

As noted briefly in Chapter 8, some categorical variables include some overlap among certain categories. Under such circumstances, the simplest alternative is to create additional categories that consist of combinations of overlapping categories to maintain the mutually exclusive

[10]Researchers should be cautious in their conclusions and clearly label any results as exploratory when post hoc model modifications have been made. Section 4.5.2 discusses this issue in another context (see also Diaconis, 1985, for full discussion of inference in exploratory data analysis).

nature of the categories comprising the nominal scale. In some cases there may be a basis in the substantive area for assigning the invidividuals with overlapping categories to a single category involved in the overlap.

An illustration of this problem arose in the 2000 U.S. census data. For the first time participants could identify themselves as a member of more than one ethnic or racial group. Investigators of population-based associations using the census data need to decide how to treat these data. Depending on their purposes and the size and location of the sample they are investigating they may try to treat each group as if it were distinct, so that each effect will be estimated with partialing for the overlap of other group membership from the variable in question. Of course, under these conditions the coding schemes for interactions cannot be employed in a straightforward manner. Alternatively, if main effects or interactions with specific combinations of ethnicity or race are anticipated, separate variables will need to be created.

No general solution to this problem can be proposed; indeed, we believe that considerable development of statistical models that address overlapping categories is likely to take place during the coming decade. The purpose of the current discussion is to remind the researcher that such issues deserve thoughtful attention. The method of coping with this fairly common problem may make a substantive difference, and should be considered carefully in the context of the purpose of the analyses.

9.2.4 Consideration of *pr*, β, and Variance Proportions for Nominal Scale Interaction Variables

In this chapter we have emphasized the cell means and differences between mean differences as represented in the study's B_i. As always, it is also possible to examine the various standardized coefficients and the proportions of Y variance that are represented by their squared values. As we first saw in Chapter 8, these coefficients are as much affected by the proportion of the sample that is in the cell as by the mean differences compared in the coefficients. Consequently, such coefficients will not generally be of as central an interest as they often are for continuous scales.

βs are usually not a focus of attention for nominal scales because the standardization (creation of a unit variance) for variables created by any of the coding methods usually makes no particular sense. If the investigator wishes to determine the unique contribution of an individual variable, it will be convenient to do so by squaring the value computed from (Eq. 9.1.1), based on β, whenever such information is not part of the computer output.

9.2.5 Summary of Issues and Recommendations for Interactions Among Nominal Scales

Many of the details of interactions among nominal scales involve careful attention to alternative coding choices, cell sizes, and study purposes. These issues will necessarily need careful consideration once the context and purposes of the analysis are clearly before the investigator. However, it may be useful to recap some of the major points made in this presentation of interactions among nominal scales.

1. All methods of coding nominal scales will yield the same aggregate significance test values for interaction sets (and, necessarily, R^2 and increments to R^2), when a full model is compared to a reduced model.

2. The coding method that is equivalent to the method used for ANOVA or ANCOVA is unweighted effects coding that treats group means as equally important, regardless of cell n. Such programs also use the residual MS from the full model to test all effects.

3. Specific B_i in different coding systems, however, will not in general have the same values, test the same mean differences or differences between differences, or have equivalent statistical significance tests. The interpretation of the unstandardized regression coefficients depends on the coding system that is used for each nominal IV and the specific interaction (or other higher order) terms, if any, that are included in the model.

4. When interaction terms are created as products of nominal scale codes for certain coding methods, these products will produce B_is corresponding to each code variable with fairly straightforward interpretations. Dummy codes, unweighted effects codes, and contrast codes produce B_is that are not affected by the sample size in each group. In contrast, weighted effect codes produce B_is that are direct functions of the relative proportion of cases in each group. Interactions in weighted effects code models can become more difficult to interpret in studies in which cell ns are not proportional (see Section 9.1.2).

5. Once again, we caution that the dummy coding option that is so often considered the "default" will frequently not be the optimal coding scheme. The coding scheme selected for each nominal IV should be the one that most directly reflects the study's hypotheses. Indeed, this is the main advantage of using a regression analysis rather than an ANOVA approach when all study factors are categorical. These are often, but not always, created as contrast codes, which have two other advantages: They are typically easy to interpret and uncorrelated in the variables except for correlations due to unequal sample sizes in the cells.

9.3 NOMINAL SCALE BY CONTINUOUS VARIABLE INTERACTIONS

9.3.1 A Reminder on Centering

As noted in all previous presentations of interactions, we will generally find that the interpretation of equations involving product terms will be more readily accomplished if the continuous variables have been centered. In the sections that follow we will assume that this centering has been done but will also note the consequences of not having done so. There is still complexity of interpretation when the continuous variable is centered, but the categorical variable(s) are not. See Section 7.2.8 for an example of a regression with interactions in which some of the IVs are centered but others are not.

9.3.2 Interactions of a Continuous Variable With Dummy-Variable Coded Groups

As is the case with equations with interactions between continuous scales, interactions between continuous and nominal scales generally employ variables that are products of the original "main effect" variables. When the nominal scale has been coded with dummy variables, these product variables will constitute a set of $g - 1$ variables that are each equal to the continuous scale for one group and zero for the other groups. When entered simultaneously with the original variables these interaction variables each reflect and test the difference between the slope for the group with a nonzero value and the reference group. Thus, the interaction between a continuous variable Z and a nominal scale W consisting of five groups coded as four dummy variables will be fully represented by an interaction set $Z \times W$ consisting of the products of each of the four dummy variables with Z.

TABLE 9.3.1
Characteristics of Separate and Combined Departments in the Academic Salary Example

Department	n	$M_{\text{Publications}}$	M_{Salary}	B_{SP}	SE_B
Psychology	60	19	\$61,719	\$1,373	\$222
Sociology	44	15.2	66,523	258	481
History	46	11.2	64,937	412	400
Mean of means		15.1	64,393	681	211
Combined	150	15.5	64,115	926	179

As an illustration let us return to our fictitious example of academic salary. In the data presented in Table 9.3.1 we have drawn a new sample from three departments in our university: psychology, sociology, and history. The primary purpose of this investigation is to see whether publications have an equivalent influence on salary in these departments. We have decided that psychology will be the reference department and have created dummy variables accordingly. Our original investigation of the main effects of the study variables is given in Table 9.3.2. In this equation the intercept is the estimated mean for a faculty member in psychology with an average number of publications (where the average is computed across the entire sample = 15.5 publications). We note that across the three departments the average return per publication is \$926. Although the observed mean salaries were not significantly different (psychology − sociology = −\$4,805 and psychology − history = −\$3,219, $t = -1.42$ and $-.90$ respectively) the department salaries adjusted for the number of publications were substantially and statistically significantly different. The estimated salary for a

CH09EX04

TABLE 9.3.2
Dummy-Variable Interactions for the Academic Salary Example

	B	SE_B	Tolerance	t	p
Main effects					
B_0	\$58,485	\$2,168			
B_P	926	193	.81	4.78	<.01
B_{D_1}	8,282	3,249	.78	2.55	<.02
B_{D_2}	10,447	3,472	.66	3.01	<.01
Regression MS = 215,311,000; $F = 8.46$; $p < .01$; $R^2 = .148$.					
Residual MS = 25,453,100.					
Full model					
B_0	\$56,922	\$2,207			
B_P	1,373	252	.46	5.44	<.01
B_{D_1}	9,669	3,235	.76	2.99	<.01
B_{D_2}	9,793	3,616	.59	2.71	<.01
B_{PD_1}	−1,115	495	.74	2.25	<.05
B_{PD_2}	−961	466	.52	2.06	<.05
Regression MS = 1,650,940,000; $F = 6.72$; $p < .01$; $R^2 = .189$.					
Residual MS = 24,559,800.					
Increment to regression MS = 89,768,000; $F = 3.66_{2,144}$; Increment to $R^2 = .041$; $p < .05$.					

psychologist with 15.5 publications is \$58,485, the intercept. Sociologists earned \$8,282 more than psychologists, considering the number of publications, and historians earned \$10,447 more, after a comparable adjustment (that is, estimated at 15.5 publications). This increase in the mean differences should not surprise us, because we saw already in Table 9.3.1 that the psychologists published more than the other disciplines. This adjustment, however, presumes a comparable "reward" per publication in the three departments, and it is this presumption that our test of the publication by department interactions is designed to test.

In the next equation we have added the two interaction terms. Once again, our intercept is the estimated effect for a psychology faculty member with an average number of publications. However, in this case the interaction terms provide for the possibility of a different effect of publications in each department. Therefore, the intercept is now the estimated salary in the psychology department for a member with 15.5 publications, taking into consideration the reward per publication in the psychology department. The regression coefficient for the "main effect" of publications is the B for publications in the psychology department (\$1,373, compare with Table 9.3.1). The B_is for the two dummy variables are the differences in salary between the psychology department and the other departments (\$66,592 – \$56,922 = \$9,669, and \$66,715 – \$56,922 = \$9,793). The regression coefficients for the two interaction terms are the differences between the publication slope (the increase in salary per publication) for each department and that for the psychology department (\$258 – \$1,373 = –\$1,115 and \$412 – \$1,373 = –\$961). Each of these differences is significant in the full model, as is the effect of publications in the psychology department. The slopes predicting salary from publications in each of the three departments are shown in Fig. 9.3.1.

Thus we see that the interaction effects for dummy variables contrast the other groups with the selected reference group with regard to both the means on the dependent variable and the slopes—the regression effects—of the continuous variable, and that the resulting coefficients are equivalent to those of the original separate groups. If other variables (U) are added to the equation without including their potential interaction effects with the nominal scale, the

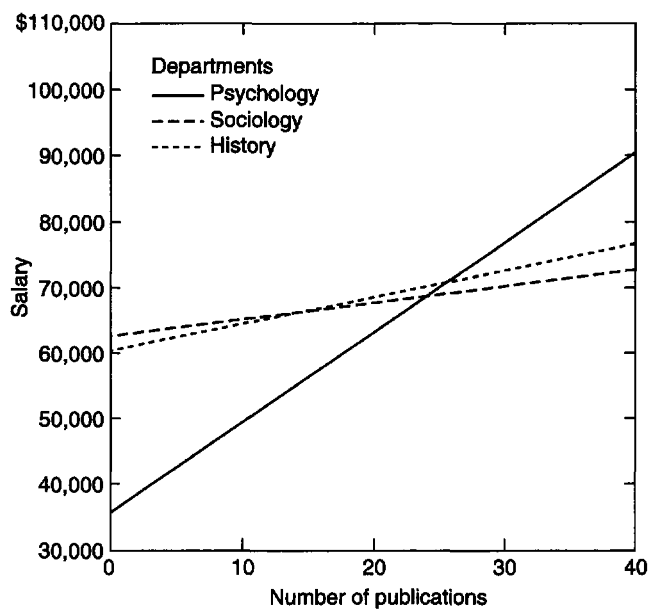

FIGURE 9.3.1 Slopes of salary on publications for three departments.

resulting B_is will vary somewhat from those that would have been obtained had the groups been analyzed separately. This is because the regression estimates treat the effects (slopes) of these other variables in every group as equivalent to those in the total group.[11]

If the continuous variable had not been centered prior to entrance into these equations, the intercept B_0 would no longer have represented the salary mean for the reference group, but rather the estimated salary for a member of that department with zero publications (which in this example is outside the range of the scores). The effects of the two dummy variables representing the other departments would similarly have represented the difference between the estimated difference between a member of each department with zero publications and a member of the psychology department also with zero publications. The interaction terms would have remained the same.

9.3.3 Interactions Using Weighted or Unweighted Effects Codes

Suppose we used effects codes to represent a nominal scale and its interaction with a continuous measure. As is equally true of the selection of the reference group in dummy variable coding, it can make a great deal of difference which group is selected to be the base group in effects codes. In this case our selection of this method of coding and of the psychology department as the base group suggests that our real interest lies in the comparison of the sociology and history departments with the "average" department. [The comparison of psychology with the average department is measured by the value $-(B_{E_1} + B_{E_2})$.]

The interpretation of B_0 in equations with effects codes is the unweighted mean of means when the continuous scales have been centered. Table 9.3.3 provides the regression analyses

TABLE 9.3.3
Effects-Coded Variable Interactions for the Academic Salary Example

	B	SE_B	Tolerance	t	p
Main effects					
B_0	\$64,728	\$1,316			
B_P	926	193	.81	4.78	<.01
B_{E_1}	2,038	1,912	.68	1.07	NS
B_{E_2}	4,204	2,039	.58	2.06	<.05
Regression MS = 215,311,000; $F = 8.46$; $R^2 = .148$.					
Residual MS = 25,453,100.					
Full model with interaction					
B_0	\$63,410	\$1,440			
B_P	681	210	.66	3.23	<.01
B_{E_1}	3,182	1,985	.61	1.60	NS
B_{E_2}	3,306	2,193	.49	1.51	NS
B_{PE_1}	423	324	.42	1.31	NS
B_{PE_2}	269	309	.43	.87	NS
Regression MS = 1,650,940,000; $F = 6.72$; $R^2 = .189$; $p < .01$.					
Residual MS = 24,559,800.					
Increment to regression MS = 89,768,000; $F = 3.66_{2,144}$; Increment to $R^2 = .041$; $p < .05$.					

[11]The reader may explore the consequences for this fictitious example by including in equations the variable time since Ph.D., which is included in the data file.

for the main effects and full model for the academic salary example. Once again, the intercept provides our estimated adjusted mean of mean salaries. In the first equation this value is slightly higher than the observed mean of means ($64,393, see Table 9.3.1) because we have used a single combined estimate of the influence of publications, $926 per publication, as estimated in the dummy variable equation as well. We note also that the adjusted mean difference between the sociology department and the adjusted mean of department means is not statistically significant, but the history department is significantly different than the average department ($p < .05$). Nevertheless, as is always true of alternative methods of coding nominal scales (Section 6.1.2), the multiple R^2 and its significance test is precisely the same as for the equation employing the dummy variable coded set.

In the full model using the effects coded scale, the adjusted mean of means is now estimated at $63,410. The B for publication is the (unweighted) mean of the department Bs for the effects of publications. The Bs for the individual effects are the adjusted mean differences between the sociology and history departments, respectively, and the adjusted mean of mean salary. And the interaction variable effects reflect the differences between the effect of publications in each of these two departments and the mean of the department mean effect of publications; that is, $258 − $681 = −$423, and $412 − $681 = $269, respectively.

Perhaps one of the most important things to note in this table is that none of the effects in the full model are statistically significant except the intercept and the mean of the mean effect of publications. *Nevertheless*, these variables fully represent the model, as can be seen by the fact that the R^2 and its significance test are identical with that of the dummy variable model for which the two interaction terms and their joint contribution added significantly. The reason for this apparent paradox may be viewed from either of two perspectives. First, we may note substantively that we decided to use the psychology department as the base department, and that this department was the most discrepant from the mean of means. Indeed, if we had used one of the other departments as the reference in the dummy-variable equation we would also have found some variables to predict less than required for statistical significance. Second, we may note that our terms in the effects coded model were less independent, that is, had lower tolerances. This lack of independence in the terms also reflects overlaps among the tests and is another way of understanding the lack of statistical significance for the individual tests in the full model.

Perhaps the most important lesson to be learned is that it is quite possible to have a *set* of variables add significantly to the prediction of a DV although none of the individual variables adds a significant unique effect. Thus, an investigator must beware of dismissing a set as of trivial importance and nonsignificant on the basis of the significance of the individual variables alone. Attention to the tolerance values usually included in the output will alert the investigator to potential problems with IV collinearity that may produce the problem we see here.[12]

9.3.4 Interactions With a Contrast-Coded Nominal Scale

Another method that is often useful for coding nominal scales is contrast coding. In this case the investigator selects orthogonal group contrasts that are of particular interest and codes them in accordance with the rules previously noted (Sections 8.5 and 9.1). Suppose our real interest was in whether the psychology department was different from the other departments either in its (adjusted) mean salary or in the influence of publications on salary. We therefore code the first nominal scale to reflect this interest (⅔ or .667 for the psychology department and −⅓ or .333 for each of the other two departments). We code the second nominal scale variable with orthogonal codes of 0 for the psychology department and .5 and −.5 for the other

[12]See further discussion of this issue in Chapter 10.

TABLE 9.3.4
Contrast-Coded Variable Interactions With a Continuous Scale for the Academic Salary Example

	B	SE_B	Tolerance	t	p
Main effects					
B_0	$64,728	$1,317			
B_P	926	193	.81	4.78	<.01
B_{C_1}	−6,243	1,923	.85	3.25	<.01
B_{C_2}	2,165	3,455	.95	.63	NS
Regression MS = 215,311,000; $F = 8.46$; $R^2 = .148$; $p < .01$.					
Residual MS = 25,453,100.					
Full model with interaction					
B_0	$63,410	$1,440			
B_P	681	211	.66	3.23	<.01
B_{C_1}	−6,487	1,923	.82	3.37	<.01
B_{C_2}	124	3,714	.79	.03	NS
B_{PC_1}	1,038	256	.77	2.70	<.01
B_{PC_2}	154	579	.82	.27	NS
Regression MS = 1,650,940,000; $F = 6.72$; $R^2 = .189$; $p < .01$.					
Residual MS = 24,559,800.					
Increment to regression MS = 89,768,000; $F = 3.66_{2,144}$; Increment to $R^2 = .041$; $p < .05$.					

two departments, respectively. The interaction terms are products of these two terms and the centered publication variable.

In Table 9.3.4 we see that in the main effects equation the intercept and B_P are the same as in the previous two models, and the effects for the contrast are the differences between the psychology department and the mean of the other two department means ($t = 3.2, p < .01$), and the difference in adjusted mean between the sociology and history departments (which also equals the difference between the two nominal scale Bs in each of the two previous main effects models using dummy or effects codes), not statistically significant. Of course, the full main effects model is precisely as large and statistically significant as it was with the previous two methods of coding.

When we move to the interaction of publications with the contrast-coded nominal scale we again find that the first contrast, comparing the psychology department with the mean of the other two departments, is statistically significant. Its value, $1,038, reflects the difference between the effect of publications in the psychology department, $1,373, and that of the mean of the other two [($258 + $412)/2 = 335]. The second contrast, between the effects of publications on salary in the sociology department as compared to the history department, is small (as we saw in the subgroup analyses reflected in Table 9.3.1) and not significant. In aggregate, again, the interaction set adds precisely the same contribution to R^2 for this coding as for any other.

9.3.5 Interactions Coded to Estimate Simple Slopes of Groups

It is not unusual for an investigator to be as interested in whether a particular variable is or is not a significant predictor of Y in each and every group. The answer to this question can be obtained in a number of ways, but perhaps the simplest one takes advantage of the fact that there are g

variables involved in the interactions: the continuous scale and the $g - 1$ variables representing the nominal scale.[13] In the previous coding methods we included the scaled variable Z and $g - 1$ main effect variables for the nominal scale and $g - 1$ interaction products. To obtain separate group slopes (simple slopes for groups) we now instead create g variables in which each group's Z values (preferably but not necessarily centered) are coded on a variable for which all other groups are coded 0. These variables are then entered simultaneously with any of the other methods of coding the $g - 1$ nominal scale variables (but with the "main effect" Z omitted). The B coefficients in this equation will be the slopes of Y on the continuous scale for each of the g groups with their appropriate standard errors and statistical significance tests.

For example, let us return to our academic department model. In Table 9.3.5 we reproduce the original dummy-variable main effects model. In the full model, however, we have removed the publications variable, and instead have included three variables representing the publications effect for each of the three departments. As noted, the first variable consists of the publications for each member of the psychology department and zero for each member of the other departments. The second variable consists of the publications for each member of the sociology department and zero for each member of the other departments. The third variable consists of the publications of each member of the history department, etc. When considered simultaneously these variables reflect precisely the slopes of the individual groups. Figure 9.3.1 presents the slopes of salary on publications for each of the three departments. The significance tests suggest that although publications have a powerful influence on salary in the psychology department, they have little or no impact on salary in the sociology or history departments.

TABLE 9.3.5
Interactions Coded for Individual Department Simple Slopes of Salary on Publications

CH09EX05

	B	SE_B	Tolerance	t	p
Main effects					
B_0	\$58,485	\$2,168			
B_P	926	193	.81	4.78	<.01
B_{D_1}	8,282	3,249	.78	2.55	<.02
B_{D_2}	10,447	3,472	.66	3.01	<.01
Regression MS = 215,311,000; $F = 8.46$; $R^2 = .148$.					
Residual MS = 25,453,100.					
Full model with interaction					
B_0	\$56,922	\$2,207			
B_{D_1}	9,669	3,235	.76	2.99	<.01
B_{D_2}	9,793	3,616	.59	2.71	<.01
B_{SD_0}	1,373	252	.90	5.4	<.01
B_{SD_1}	258	426	1.00	.60	NS
B_{SD_2}	412	392	.73	1.05	NS
Regression MS = 1,650,940,000; $F = 6.72$; $R^2 = .189$; $p < .01$.					
Residual MS = 24,559,800.					
Increment to regression MS = 89,768,000; $F = 3.66_{2,144}$; Increment to $R^2 = .04$; $p < .05$.					

[13]Aiken and West (1991) provide a method for hand calculation of individual slopes, and note that one can also alternate the reference group in a dummy variable model, since the coefficient for the continuous variable in that model and its significance test represent the simple slope for the reference group.

(The reader is reminded that any resemblance of these data to actual departments is purely coincidental.) These analyses may be repeated using contrast or effects codes to confirm that the slope estimates are not affected by the method of coding the main effects of the nominal scale.

Perhaps the most crucial thing for the reader to note is that all of these methods share precisely the same significance test values for the total contribution of the interaction set to the prediction of Y. However, as noted, depending on the coding selected, the individual variables may all be statistically significant or none of them may be individually statistically significant. Thus it is important to check on the contribution of the set as a whole, especially if one is concerned about the possible significance of the set as a check on the assumption of equal slopes. And it is also important to select a coding method that is consistent with the investigator's need to detect and estimate the most theoretically or practically important interactions.

We also note that although the equations we have examined here using any of the coding methods produced coefficients that were relatively simple functions of the raw coefficients in subgroups, that is by no means necessarily the case in practice. This is because typically the researcher will include one or more other variables in the equation for which interaction terms are not hypothesized and, often, not investigated. The slopes for these variables are thus assumed to be equivalent for subgroups, although they are hardly ever precisely so. The effects of these covariates on the adjusted estimates of group by continuous scale interactions will be based on the covariates' full sample average effects, and may well influence the Bs and significance tests of the interaction variables (West, Aiken, and Krull, 1996).

Examination of Group Differences at a Point Other Than the Mean of the Continuous Scale

It is sometimes of particular interest whether the group means are significantly different at a particular point on the continuous scale. For example, in our academic example a member of the sociology department counters the findings by saying that any departmental differences in rewards for publications do not have effects on the more accomplished (that is, published) members of the department. He defines this group as those with 20 publications.

To test this hypothesis we begin by noting that the comparison of Y means of the groups in the full model is at a score of zero on the continuous variable. This point equals the full sample continuous variable mean when the continuous scale has been centered. Thus, by simple extension we may subtract from our continuous scale scores any other constant in order to make zero the point of interest and proceed with the analysis, as before. Regression coefficients reflecting mean differences among cells will now reflect those Y differences *at the selected point.*

Therefore, we rescore publications by subtracting from the original scores 20 rather than the mean. Running the full equation with dummy variables representing the comparison of
CH09EX06 the sociology and history departments with psychology, including interactions, we find the estimated effects as shown in Table 9.3.6.

In the first model we reproduce the equation estimates from the dummy-variable model using the centered publication variable. As noted earlier, the D_1 and D_2 variables show significant t values, indicating that both the sociology and history departments have a higher estimated salary for those members with the mean (aggregated across department) number of publications, taking into account the differences in the publication effects in the different departments.

The next column shows the equation estimates when the publications variable has been "re-centered" by subtracting 20 from each value, rather than the 15.49 publications subtracted

TABLE 9.3.6
Estimation of Department Salary Differences at Alternative Levels of Publications

IV	Predictor: Publications centered at mean B(SE)	t	Predictor: Publications – 20 B(SE)	t	Predictor: Publications – 10 B(SE)	t
Intercept	\$56,922 (\$2,207)		\$63,114 (\$2,039)		\$49,385 (\$3,030)	
D_1	\$9,669 (\$3,235)	2.99	\$4,640 (\$3,726)	1.25	\$15,790 (\$4,448)	3.55
D_2	\$9,793 (\$3,616)	2.71	\$5,458 (\$4,633)	1.18	\$15,069 (\$3,846)	3.92
Publications	\$1,373 (\$252)	5.44	\$1,373 (\$252)	5.44	\$1,373 (\$252)	5.44
$D_1 \times$ *Publications*	−\$1,115 (\$495)	2.26	−\$1,115 (\$495)	2.26	−\$1,115 (\$495)	2.26
$D_2 \times$ *Publications*	−\$961 (\$466)	2.06	−\$961 (\$466)	2.06	−\$961 (\$466)	2.06

when we centered in the usual way. It turns out that our sociologist was right (in this fictitious example), and the estimated differences between department salaries for members with 20 publications is not statistically significant. We note, however, that the interactive effects, reflecting the differences in the effect of publications on salary in the different departments remain unaltered.

"Aha!" counters a junior member of the psychology department. "That's all very well for senior members, but the situation is even worse for those of us who have not yet had time to publish. Look at members with 10 publications in each department."

In the final set of columns in Table 9.3.6 we present the same equation in which 10 is now subtracted from each person's publications. As suspected by our junior colleague, the estimated mean difference in department salary for those with 10 publications is over \$15,000 (taking into account the differences in the effects of publications). Necessarily, in all of these equations the different linear transform of publications has not altered the estimates of the main effects or interactions of publications, since the slopes cannot be affected by subtracting a constant from a variable. Nor can there be any influence of this transform on the overall R^2 or its statistical significance.

9.3.6 Categorical Variable Interactions With Nonlinear Effects of Scaled Independent Variables

Just as nominal scale interactions with a continuous scale are readily accomplished, nominal scale interactions with additional functions or powers of a continuous scale are also entirely feasible. The constraints against routine inclusion of such interaction terms are more likely to come from an inadequacy of theory to predict their presence and meaning than any difficulties in computing their effects. And, as we noted in Chapter 6, any problems of unreliability will tend to diminish our power to detect curvilinearity in relationships even more than it does for linear relationships. When the effective sample size is cut by examination of the subgroups represented by a nominal scale, such statistical power considerations are even more restricting.

Thus, although we present these methods here, the investigator should reflect carefully on the theory that predicts such effects, on the expected size of the difference, and on the statistical power considerations that may restrict one's ability to detect differences in curve shapes in subgroups.

The method is a straightforward extension of the previously presented treatment of interactions with nominal scales. One may use any of the methods of coding the nominal scale, as appropriate to the researcher's purpose and hypotheses. Interaction terms with continuous scales, including power functions, are created as simple products of the main effect sets. Thus, with 4 groups and z (centered, of course) and z^2, there will be six interaction terms created by the product of the $g - 1 = 3$ nominal scale variables and the two z variables.

For example, suppose that in our academic department salary investigation we had a further hypothesis. This hypothesis posits that these departments had different policies with regard to the influence of seniority on salary. It is believed that the history department is inclined to value the perspective and prestige associated with long familiarity with and by the professional field, and thus salary increases may be enhanced by seniority, perhaps even more than linearly. In psychology, it is hypothesized, academics tend to forget research and theoretical contributions that are more than a few years old, and thus the influence of seniority on salary diminishes over time. No specific hypothesis is made about the effects of seniority in the sociology department.

The analyses of these effects in our fictitious example are presented in Table 9.3.7 for each of the individual departments.[14] As can be seen, the linear effect of seniority is largest in the CH09EX07 psychology department, but there is also a significant negative curvilinear effect, indicating an overall "frown" effect—a downturn from the linear trend at the upper end (as hypothesized).

TABLE 9.3.7
Differential Curvilinear Effects of Seniority in the Academic Department Example: Individual Department Effects

	Coefficient value	SE	t (or F)
Psychology department			
B_0	\$65,670	\$2,548	25.8
$\text{Time}_c = B_1$	\$2,424	\$390	6.2; $df = 57$
$(\text{Time}_c)^2 = \text{B}_2$	−\$143	\$49	2.9; $df = 57$
Total R^2	.404		$F = 19.3$; $df = 2, 57$
Sociology department			
B_0	\$63,319	\$3,677	17.2; $df = 41$
$\text{Time}_c = B_1$	\$930	\$539	1.7; $df = 41$
$(\text{Time}_c)^2 = B_2$	\$109	\$105	1.0; $df = 41$
Total R^2	.112		$F = 2.58$; $df = 2, 41$
History department			
B_0	\$64,535	\$2,820	22.9; $df = 43$
$\text{Time}_c = B_1$	\$1,606	\$492	3.3; $df = 43$
$(\text{Time}_c)^2 = B_2$	−\$61	\$78	.8; $df = 43$
Total R^2	.210		$F = 5.72$; $df = 2, 43$

[14] In these analyses we have centered years at the combined department mean, which makes the intercepts somewhat different than they would be if each department's years were centered separately but also makes the findings more directly comparable with the combined analysis. Also note that the dependent variable salary is not the same one used in the previous section.

Neither the linear nor the quadratic trend is statistically significant in the sociology department. In the history department there is a statistically significant linear trend of $1,606 per additional year, but no significant quadratic trend.

Analyzing these three departments simultaneously in order to test the significance of the curvilinear component, we will use dummy-variable coding with the psychology department as the reference group. There will be eight IVs in these analyses, two variables representing the main effects for groups, centered years and centered years squared, the two products of the group variables with years, and the two products of the group variables with years squared. The hierarchical regression analyses are presented in Table 9.3.8, where we can see that there was a generally significant quadratic effect, and that the quadratic effect differed among the departments, although not precisely as anticipated. The quadratic term was $252 less negative in the sociology department than in the psychology department (noting that the first dummy variable compares the sociology department to the psychology department), and only $83 less negative (and not significantly so) in the history department. Note that these are precisely the differences that we noted in the analyses of the individual departments where the sociology = psychology difference = $109 − (−$143) = $252 and the history psychology difference = −$61 − (−$143) = $83 (within rounding error). Now we also know that the psychology department is more likely to show a decline in seniority effects on salary among its most senior members to a greater extent than is the sociology department (which in fact is hardly influenced by seniority at all).

Once again, we can usefully reconstruct the slopes of each department in our combined analyses by using our group slope coding method in which we omit the continuous variable(s) and include a set of variables in which each group's (centered) continuous variable is coded on

TABLE 9.3.8
Hierarchical Multiple Regression Analysis of Seniority Effects

Model	$B(SE)$		Tolerance	t	Increment to R^2, F, numerator df
Main effects					.236; 15.0; 3 df
Intercept	$60,370	($2,058)		29.3	
Department dummy 1	$5,487	($3,153)	.81	1.7	
Department dummy 2	$2,653	($3,152)	.79	0.8	
Time_c	$1,548	($251)	.96	6.2	
Main effects + Time_c^2					.016; 3.0; 1 df
Time_c^2	−$69	($40)	.84	1.7	
Main effects, $\text{Time}_c^2 + T \times D$					.015; 3.0; 2 df
$T \times \text{Dept}_1$	−$1,050	($612)	.60	1.7	
$T \times \text{Dept}_2$	−$545	($599)	.58	0.9	
Main effects, $T \times D, T^2 \times D$					.028; 7.6; 2 df
Intercept	$65,670	($2,689)		24.4	
Dept_1	$ 2,424	($412)	.34	5.9	
Dept_2	−$2,351	($4,239)	.43	.6	
Time_c	−$1,135	($4,022)	.46	.3	
Time_c^2	−$143	($52)	.49	2.8	
$\text{Dept}_1 \times \text{Time}_c$	−$1,494	($633)	.54	2.4	
$\text{Dept}_2 \times \text{Time}_c$	−$818	($665)	.46	1.2	
$\text{Dept}_1 \times \text{Time}_c^2$	$252	($107)	.43	2.4	
$\text{Dept}_2 \times \text{Time}_c^2$	$83	($97)	.37	.4	

TABLE 9.3.9
Quadratic Interaction Example Coded for Simple Quadratic Slopes

Predictor	$B(SE)$		Tolerance	t	*Cumulative* R^2; F; df
Intercept ($= M_M$)	\$64,507	(\$1,730)		37.3	
$Psychology_{(\text{vs. others})}$	\$1,743	(\$3,486)	.54	.5	
$Sociology_{(\text{vs. history})}$	−\$1,216	(\$4,437)	.54	.3	.014; $F = 1.1$; 2, 147 *df*
$Time_{c(\text{psychology})}$	\$2,424	(\$412)	.80	5.9	
$Time_{c(\text{sociology})}$	\$930	(\$481)	.94	1.9	
$Time_{c(\text{history})}$	\$1,606	(\$522)	.74	3.1	.189; $F = 6.7$; 5,144 *df*
$Time^2_{c(\text{psychology})}$	−\$143	(\$52)	.63	2.8	
$Time^2_{c(\text{sociology})}$	\$109	(\$94)	.56	1.2	
$Time^2_{c(\text{history})}$	−\$61	(\$83)	.52	.7	.295; $F = 7.4$; 8,141 *df*

a variable on which the other groups are coded zero. When we wish to examine the quadratic effects of the individual groups in the simultaneous analysis, we simply square the group slope variables. As was true in the example presented in the previous section, the increments to R^2 will be the same as in the previous hierarchical analysis (except the quadratic term of the continuous variable will be included with the final step rather than earlier).

In Table 9.3.9 the final equation is presented for the academic salary example in which we have used contrast codes (contrasting psychology with the other two departments and sociology with history) for the group main effects.[15] These linear and quadratic slope components for the individual departments are precisely what we obtained in their separate analyses in Table 9.3.6, and the R^2 for the full equation is precisely what it was for the previous analysis using dummy codes in the interactions.

The standard errors of the Bs are not the same as in the individual department analyses, since they are based on the full combined sample n. We note that neither the linear nor the quadratic effect was statistically significant in the sociology department. As is always the case, these slopes should be graphed in order to make sure that the interpretation is correct. Figure 9.3.2 presents the quadratic slopes for the three departments. Here we see clearly that in the psychology department the linear effect of seniority is steepest, but also the "fall off" in salary at the upper end is greatest.

Of course, if one only needed to know the shape of the quadratic slopes for each of these groups one could just as well run the equations separately for the departments, as we did in Table 9.3.7. However, when there are other predictors in the model, for which one is either ready to assume that the effects are roughly equivalent across groups, the adjusted group linear and quadratic slopes can most readily be obtained in the combined sample.

9.3.7 Interactions of a Scale With Two or More Categorical Variables

Just as easily as one can accomplish an examination of possible group differences in the effects of a continuous variable with one nominal scale, one may include interactions with two or more such scales in one's equations. In fact, it is not at all unusual to do so when a scale has only two categories; e.g., sex, handedness, treatment, residence type, zygosity, and many demographic variables often appear as dichotomies in studies. Interactions with more than one

[15] The reader is reminded that it doesn't matter for any of the simple slope variables whether we use dummy, effects, or contrast codes for the main effects of groups.

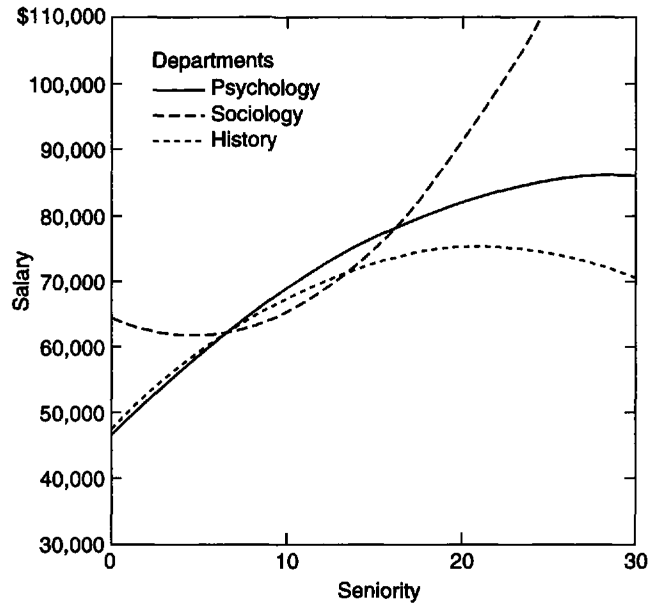

FIGURE 9.3.2 Quadratic slopes of salary on seniority.

such two-group variable are often plausible and of interest. When the sample is divided into more groups, however, the statistical power to detect multiple interactions is likely to quickly dissipate. Thus investigators are likely to need to exert considerable substantive and statistical judgment in the decisions as to which interactions to include.

Interactions involving continuous scales with multiple nominal scales also require consideration of whether the interaction terms *among the nominal scales* need to be included in the interactions with the continuous scales. As we noted in Section 9.2.1, with three nominal scales, even when two of these were dichotomies and the other only 3 groups, there are, in addition to four main effects, seven interaction terms. Thus, the full representation of the interactions of these groups with a continuous scale would require 11 interaction products of the nominal scale codes with the continuous scale. Unless the sample is huge, or the expected differences very large indeed, the power of the aggregate contribution to R^2 is likely to be small. As we have noted in Chapter 5, it is often a useful strategy to protect oneself against an excessive number of Type I errors ("findings" that are not characteristic of the population and thus will not replicate) by testing the contribution of a set of variables to R^2 for statistical significance before proceeding to examine the effects of individual variables. As always, such protection comes at a price, in terms of the risk of Type II errors (failure to detect effects that are actually present in the population). The investigator's goal must always be to balance these risks in ways that will further the scientific contribution of the study.

One possibility to be considered when the potential number of interaction terms in the equation mushrooms, is to examine only some of the potential group differences. For example, in the contrast-coded example of the effect of publication on salary presented in Section 9.3.4 our hypothesis and interest was in the comparison of the psychology department with the other two departments. The second contrast, that between the other two departments, was present only for completeness. This second term could have been omitted altogether (and its interaction term as well).

As we have noted repeatedly, beginning in Section 8.1.1, when a nominal scale is represented by $g - 1$ variables, these variables considered together may reflect different differences than they do when each is considered separately. In particular, although each variable in a dummy-coded set reflects a dichotomy between one group and all other groups, when considered simultaneously in a regression equation these individual effects are comparisons between that group and the group consistently coded zero, the reference group. This means that consideration of omitting some of the variables in a nominal scale set, whether in the main effects or in interaction terms involving the variable, needs to be done with great care to make sure that the remaining variables accurately reflect the comparison of interest. The easiest way to ensure selection appropriateness is by contrast coding for the desired effect.

For example, in the previous section we determined the difference in curvilinearity of the seniority effect on salary by interactions with the dummy-variable coded academic departments. However, as stated, our hypothesis had only to do with a difference in curvilinearity between the psychology and history departments, with no prediction made for the sociology department. For the dummy variable interaction we used four variables, the product of the two dummy variables reflecting the comparison of the sociology and history departments, respectively, with the linear and quadratic aspects of seniority (time since Ph.D.). However we might instead have contrast coded for the term that was actually of interest (1 for psychology, −1 for history, 0 for sociology). In computing the interaction effect then we could have omitted two of the interaction terms and thus increased our power to detect the remaining two, considered as a set.[16]

As noted at many points in the book, such strategic considerations are the hallmark of thoughtful research, and must be justified by the "principled argument" (Abelson, 1995) of the investigator.

9.4 SUMMARY

The chapter begins with a summary of how the interactions among nominal scales reflect differences between differences in cell means on Y. The 2 by 2 design is revisited, and cases with no interaction, crossed interaction, and uncrossed interaction are illustrated to reveal the meaning of a joint effect reflected in an interaction term. Alternative methods of coding nominal scales are reviewed. It is noted that when scales are coded by methods that produce correlations between main effects and interactions, it will generally be necessary to carry out hierarchical analyses to test the effects of the interaction set (Section 9.1.1). Analysis of k by k designs using different coding methods for the nominal scales is discussed (Section 9.1.2). Such an analysis begins by testing unequal n experimental studies using ANOVA-equivalent methods. Other research designs may employ variations of hierarchical MRC analyses to determine overall group effects and interactions. An illustrative example is presented.

In Section 9.2 interactions involving more than two nominal scales are discussed and illustrated with a 3 by 2 by 2 example. Subsections also discuss the possibilities of omitting some interaction terms and of having some overlap in the categories of one or more nominal scales. The utility of the various coefficients resulting from a regression analysis including interactions among nominal scales is reviewed (Section 9.2.4). Finally, a summary of strategic issues and recommendations for interactions among nominal scales is offered (Section 9.2.5).

Section 9.3 considers the interaction between continuous and nominal scales. After a reminder on the general utility of centering the continuous variable, an example is carried through the following section illustrating the consequences of coding with dummy variables,

---

[16]As it turned out, the significant difference was found with the sociology department, or with the combined sociology and history departments, rather than with the history department. If only the latter had been tested, this "finding" would not have emerged.

effects codes, and contrast codes. It is emphasized that the selection of the coding method should be closely linked to the investigator's hypotheses, and that the coefficients and significance tests of the individual variables are not at all equivalent across coding schemes. However, as is necessarily the case with alternative coding methods for nominal scale main effects (Chapter 8), all methods produce in aggregate the same contributions to R^2 of both main effects and interactions when analyzed hierarchically (Sections 9.3.2 to 9.3.4). Section 9.3.5 describes and illustrates an easy method of obtaining simple slopes, that is, the increase in Y per unit of the continuous variable, for the individual categories in a nominal scale. It also reviews a method for determining whether the Y means of subgroups are significantly different at a point on the continuous variable other than its mean.

Section 9.3.6 reviews the considerations that should operate when planning an analysis of the difference in curvilinearity of effect of a continuous variable for subgroups on a nominal scale. An illustration of such an analysis is provided. Finally, Section 9.3.7 discusses the interactions of a scaled variable with two or more nominal scales and reviews the considerations in determining whether and which such interactions should be included, and how codes should be selected when not all interaction terms are represented.