6

Modeling Interactions

Interactions as Conditional Differences

In additive regression models, the coefficients for dummy independent variables capture differences between groups on the dependent variable, whereas the coefficients for interval independent variables capture the additional contribution of each unit of the interval variable. Occasionally, researchers consider interactive effects between two independent variables on the dependent variable. This requires modeling interactions.

Table 6.1 shows the means for math scores. The pattern in math scores is that those with a college-graduate parent score higher than those not with a college-graduate parent and those living with two biological parents score higher than those not living with two biological parents. The general pattern for the differences on the parental education variable holds within family structure categories. At the same time, the general pattern for differences on the family structure variable also holds within categories of the parental education variable.

Table 6.2 shows differences in math scores between categories of the family structure variable within categories of the parental education variable. The two-biological-parent category is the contrast group. The means for biological-parent/stepparent category and the single-parent category are lower than the mean for the two-biological-parent

Table 6.1 Mean Math Scores by Family Structure and College-Graduate Parent

Family Structure	Not College Grad.	College Grad.	Total
Two Bio.	49.27	56.28	52.88
Bio./Step.	48.18	53.72	50.32
Single	48.19	54.44	50.14
Other Fam.	46.95	53.83	49.10
Total	48.59	55.54	51.63

Table 6.2 Differences in Mean Math Scores by Family Structure and College-Graduate Parent

Family Structure	Not College Grad.	College Grad.	Difference in Differences
Two Bio.	—	—	—
Bio./Step.	−1.09	−2.56	−1.47
Single	−1.08	−1.84	−.76
Other Fam.	−2.32	−2.45	−.13

category both among those not with a college-graduate parent and among those with a college-graduate parent.

The pattern of mean differences by family structure can be compared for the two groups by taking the difference in the differences. This calculation shows that the mean differences by family structure are more negative in the college-graduate-parent category than in the not-college-graduate-parent category. The differences in the differences show whether there is interaction. The idea of interaction, then, is simply that mean differences on the first independent variable are conditional on the second independent variable.

Another way to compare is to switch which variable is used as the "conditioning" variable. Since the parental education variable was the conditioning variable in Table 6.2, I use family structure as the conditioning variable in Table 6.3.

Table 6.3 shows that although those not with a college-graduate parent have a lower math score than those with a college-graduate parent, the difference is less in the biological-parent/stepparent and single-parent categories than in the two-biological-parent category. These differences in differences are shown in the last column.

Note that the differences in differences are the same for when I use the parental education variable as the conditioning variable as when I use the family structure

Table 6.3 Differences in Mean Math Scores by Family Structure and College-Graduate Parent

Family Structure	Not College Grad.	College Grad.	Difference in Differences
Two Bio.	—	7.01	—
Bio./Step.	—	5.54	−1.47
Single	—	6.25	−.76
Other Fam.	—	6.88	−.13

variable as the conditioning variable. Thus, the same set of differences in differences can be interpreted in two different ways depending on what variable is chosen as the conditioning variable.

Interactions Between Dummy Variables

The first step in considering interactions between dummy variables in regression is to examine the additive model shown in Equation 1. This model includes C_1, which is "1" if college graduate parent and "0" if not. The model also includes family structure variables for biological parent/stepparent, single parent, and other family. C_0, not-college-graduate parent, and F_1, two biological parents, are the excluded variables:

$$Y = a + b_1C_1 + b_2F_2 + b_3F_3 + b_4F_4 \qquad (1)$$

The following matrices show the data matrices for the additive model. The essence of an additive model is that the effects of one variable are not conditioned on the values of a second variable. We could say the values on one variable are independent of the values of the second. The coefficient for the intercept is the mean for not-college-graduate/two biological parents:

$$\begin{array}{ccccc} U & C_1 & F_2 & F_3 & F_4 \\ \begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\1\\1\\1\\1\end{bmatrix} & \begin{bmatrix}0\\1\\0\\0\\0\\1\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\1\\0\\0\\0\\1\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\1\\0\\0\\0\\1\end{bmatrix} \end{array}$$

Equation 2 includes interaction variables. We can create the interactions variable between two sets of dummy variables by multiplying all the variables in the first set by all the variables in the second set:

$$Y = a + b_1C_1 + b_2F_2 + b_3F_3 + b_4F_4 + b_5C_1F_2 + b_6C_1F_3 + b_7C_1F \qquad (2)$$

U	C_1	F_2	F_3	F_4	C_1F_2	C_1F_3	C_1F_4
1	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0
1	0	0	1	0	0	0	0
1	0	0	0	1	0	0	0
1	1	0	0	0	0	0	0
1	1	1	0	0	1	0	0
1	1	0	1	0	0	1	0
1	1	0	0	1	0	0	1

The interaction variables are nested in what we might call the additive variables. All three interaction variables are nested within C_1. In addition, C_1F_2 is nested within F_2, C_1F_3 is nested within F_3, and C_1F_4 is nested within F_4. As you can see, the interaction variables are nested in two different additive variables. The effects of family structure depend on parental education, and the effects of parental education depend on family structure. Thus, there is a decision about what is the "conditioning" variable.

The differences as captured by the interaction variables can be viewed in two ways. They can be viewed as capturing differences in parental education effects by family structure or as capturing differences in family structure effects by parental education.

In addition to creating an equation that models differences in effects for groups, we can create an equation that models the effects within each group. Equation 3 estimates the effects of family structure within categories of parental education. The change from the previous equation is that rather than using the F_2, F_3, and F_4 variables, the equation includes C_0F_2, C_0F_3, and C_0F_4:

$$Y = a + b_1C_1 + b_2C_0F_2 + b_3C_0F_3 + b_4C_0F_4 + b_5C_1F_2 + b_6C_1F_3 + b_7C_1F_4 \qquad (3)$$

Equations 2 and 3 take different approaches to modeling interactions. Equation 2 estimates the effects of family structure for those not with a college-graduate parent with the F_2, F_3, and F_4 coefficients and then the difference from those effects for those with a college-graduate parent with the C_1F_2, C_1F_3, and C_1F_4 coefficients. In contrast, Equation 3 estimates the effects of parental structure for those not with a college-graduate parent with the C_0F_2, C_0F_3, and C_0F_4 coefficients and then the effects of family structure for those with a college-graduate parent with the C_1F_2, C_1F_3, and C_1F_4 coefficients. I refer to the first equation as a “standard interaction model” and to the second equation as a “within-group effects” model.[1]

We can see the relationship of the interaction variables in the within-group effects model in the following matrices. The C_0F_2, C_0F_3, and C_0F_4 variables and the C_1F_2, C_1F_3, and C_1F_4 are not nested within one another. What is happening is the F_2 variable has been split into two parts with C_0F_2 being those in both category C_0 and category F_2 and with C_1F_2 being those in both category C_1 and category F_2. Thus, in the within-group effects model, we split the F_2 variable into two parts, one for those in C_0 and one for those in C_1. The same is true for F_3 and F_4.

The within-group effects model is the model that contains the coefficients that the standard interaction model is testing the differences between.[2] The primary reason that researchers may have difficulty interpreting interaction coefficients in the standard interaction model is a lack of clarity about the underlying within-group effects model:

U	C_1	C_0F_2	C_0F_3	C_0F_4	C_1F_2	C_1F_3	C_1F_4
1	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0
1	0	0	1	0	0	0	0
1	0	0	0	1	0	0	0
1	1	0	0	0	0	0	0
1	1	0	0	0	1	0	0
1	1	0	0	0	0	1	0
1	1	0	0	0	0	0	1

Table 6.4 shows results from regression analysis that examines the effects of family structure and parental education on math scores. Model 1 is the additive model and shows that those with a college-graduate parent score higher in math than those not with a college-graduate parent. The model also shows that those living with two biological parents score higher in math than those not living with two biological parents. An additive model assumes that the effect of family structure is the same at all levels of parental education. For example, the disadvantage of living

1. Gordon (2010), pages 253–277, presents an alternative discussion of the standard interaction model using the concept of conditional means.

2. Demaris (2004), page 147; Jaccard (1990), pages 42–45; and Hardy (1993), pages 44–46, show how to calculate within-group effects by hand but do not show how to use dummy variables to estimate those coefficients. The disadvantage of simply adding coefficients to calculate within-group effects is that no standard error for the coefficient is created.

Table 6.4 Interactions Between Family Structure and College-Graduate Parent With Math Score as Dependent Variable

	Model		
	1	2	3
Independent Variable	***B***	***B***	***B***
College-Graduate Parent	6.65*	7.01*	7.01*
Bio./Step.	−1.70*	−1.09*	—
Single	−1.39*	−1.08*	—
Other Fam.	−2.43*	−2.32*	—
Not College Grad. × Bio./Step.	—	—	−1.09*
Not College Grad. × Single	—	—	−1.08*
Not College Grad. × Other Fam.	—	—	−2.32*
College Grad. × Bio./Step.	—	−1.47*	−2.56*
College Grad. × Single	—	−.76*	−1.84*
College Grad. × Other Fam.	—	−.13	−2.45*
Intercept	49.45	49.27	49.27
R^2	.124	.125	.125

*$p < .05$.

in a single-parent family would be the same despite whether the student's parent had a college degree.

Model 2 in Table 6.4 is the standard interaction model.[3] A common way of misinterpreting the standard interaction model is to discuss the model as one would for control models. An example of such a misinterpretation would be to say "the effect for college graduate parent increased going from Model 1 to Model 2 and the effects for family structure decreased when the interactions were controlled." Although the standard interaction model is created by adding variables to the additive model, what happens when those variables are added is definitely not like what happens when variables are added in control modeling.

To understand properly what is happening when the interaction variables are added to the additive model, we should first estimate Model 3, which is the within-group effects model. Model 3 in this example takes the family structure variables in

3. Linneman (2014), Chapter 12, provides a basic introduction to calculating and interpreting interactions in regression.

Model 1 and splits them into two sets of effects, one set for those not with a college-graduate parent and one for those with a college-graduate parent. Model 3 shows that the family structure effects are more negative for those with a college-graduate parent than for those not with a college-graduate parent. What Model 3 does not tell us is whether the two sets of coefficients are significantly different from one another. If the two sets are not significantly different, then Model 1 is the correct specification. If the two sets of coefficients are significantly different, then Model 3 is the correct specification.

Thus, Model 2 tests whether the family structure effects for those not with a college-graduate parent are different from the family structure effects for those with a college-graduate parent. Two coefficients for the interaction variables in Model 2 are significantly different from zero. This shows that the effects for family structure for those with a college-graduate parent are significantly more negative than the coefficients for those not with a college-graduate parent.

So far, we have examined the effects of family structure conditioned on parental education. The second way to consider interactions is to examine the effects of parental education conditioned on family structure as shown in Equation 4 and the following matrices:

$$Y = a + b_1F_2 + b_2F_3 + b_3F_4 + b_4C_1 + b_5C_1F_2 + b_6C_1F_3 + b_7C_1F_4 \qquad (4)$$

U	F_2	F_3	F_4	C_1	C_1F_2	C_1F_3	C_1F_4
1	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0
1	0	1	0	0	0	0	0
1	0	0	1	0	0	0	0
1	0	0	0	1	0	0	0
1	1	0	0	1	1	0	0
1	0	1	0	1	0	1	0
1	0	0	1	1	0	0	1

The coefficient for C_1 in Equation 4 measures the effect of parental education for those living with two parents. The coefficients for C_1F_2, C_1F_3, and C_1F_4 measure the additional effect of parental education in the other three categories.

Equation 5 shows the within-group effects model where the parental education effect is conditioned on family structure. It uses C_1F_1 rather than C_1 as in Equation 4. The estimated coefficient for C_1 in Equation 4 and the coefficient for C_1F_1 in Equation 5 are the same, and both capture the parental education effect for those with two biological parents. However, Equation 5 estimates the parental education effect for those in each other family structure type rather than the differences for those types:

$$Y = a + b_1F_2 + b_2F_3 + b_3F_4 + b_4C_1F_1 + b_5C_1F_2 + b_6C_1F_3 + b_7C_1F_4 \qquad (5)$$

$$\begin{matrix} U & F_2 & F_3 & F_4 & C_1F_1 & C_1F_2 & C_1F_3 & C_1F_4 \\ \begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix} & \begin{bmatrix}0\\1\\0\\0\\0\\1\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\1\\0\\0\\0\\1\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\1\\0\\0\\0\\1\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\1\\0\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\0\\1\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\0\\0\\1\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\0\\0\\0\\1\end{bmatrix} \end{matrix}$$

The C_1F_1, C_1F_2, C_1F_3, and C_1F_4 variables in the previous matrices are the variables that result when the C_1 variable is subdivided into parts for each family structure category. Table 6.5 shows the results for the within-group effects model that uses the subdivided variables. In Model 3, the parental education effects appear less positive for those not living with two biological parents. The standard interaction model, Model 2, shows that the parental education effects for those not living with two biological parents are

Table 6.5 Interactions Between Family Structure and College-Graduate Parent With Math Score as Dependent Variable

Independent Variable	Model 1 *B*	Model 2 *B*	Model 3 *B*
Bio./Step.	−1.70*	−1.09*	−1.09*
Single	−1.39*	−1.08*	−1.08*
Other Fam.	−2.43*	−2.32*	−2.32*
College-Graduate Parent	6.65*	7.01*	—
Two Bio. × College Grad.	—	—	7.01*
Bio./Step. × College Grad.	—	−1.47*	5.54*
Single × College Grad.	—	−.76*	6.25*
Other Fam. × College Grad.	—	−.13	6.88*
Intercept	49.45	49.27	49.27
R^2	.124	.125	.125

*p < .05.

significantly less than the parental education effect for those living with two biological parents. The significant interactions in Model 2 indicate that Model 3 is the correct specification for parental education effects, not Model 1.

I have shown that although there is one standard interaction model for the interaction between two sets of dummy variables, there are two different within-group effects models. In the examples in Table 6.5, one within-group effects model showed that family structure effects are greater for those living with a college-educated parent than for those not. The other within-group effects model showed that the parental education effect was more for those living with two biological parents than for those not.

These are two ways of addressing the same underlying issue. To say that the disadvantage of not living with two biological parents is more for those living with a college-educated parent is the same as saying that the advantage of living with a college-educated parent is less for those not living with two biological parents. Depending on which variable that we choose as the conditioning variable, we can interpret the interaction results in two ways. I reiterate that it is imperative to decide on which variable is the conditioning variable to interpret a standard interaction model properly.

In the previous discussion, I outlined two types of interaction models, the standard interaction model and the within-group effects model. Equation 6 is a third type of model that I refer to as the "all differences" interaction model. This model estimates only first-order differences.

The parental education variable has two categories, and the family structure variable has four categories. Thus, there are eight combinations of the two variables. The all-differences interaction model estimates the difference in means for seven of the categories from one of the categories, the one chosen as the excluded variable. In Equation 6, the C_0F_1 variable is excluded.

The data matrices for the equation include seven interaction variables and the unit vector. The interaction variables are nested only within the unit vector. The coefficients for the interaction variables estimate the difference in the mean for the group captured by the interaction variable and the mean for the contrast group. The all-differences interaction model is a type of additive model because the interaction variables are not nested:

$$Y = a + b_1C_0F_2 + b_2C_0F_3 + b_3C_0F_4 + b_4C_1F_1 + b_5C_1F_2 + b_6C_1F_3 + b_7C_1F_4 \tag{6}$$

$$\begin{array}{cccccccc} U & C_0F_2 & C_0F_3 & C_0F_4 & C_1F_1 & C_1F_2 & C_1F_3 & C_1F_4 \\ \begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix} & \begin{bmatrix}0\\1\\0\\0\\0\\0\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\1\\0\\0\\0\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\1\\0\\0\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\1\\0\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\0\\1\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\0\\0\\1\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\0\\0\\0\\1\end{bmatrix} \end{array}$$

There is one more model that involves interactions between two set of dummy variables, and I refer to this model as the "all-means" model. In this example, the all-means model involves including all eight interaction variables capturing the interaction between parental education and family structure. The unit vector is not included. The coefficients in this model estimate the mean for each subgroup, hence, the name "all-means model." This is not a model that we would estimate for research purposes. However, the model is useful for purposes of understanding interactions and the role of the intercept.

When we look at the data matrices that follow Equation 7, we observe that none of the variables is nested within any other variable and that is why the model would estimate all means. Substituting the unit vector for any one of the interaction variables would lead to estimation of differences from the mean for the excluded category:

$$Y = b_1C_0F_1 + b_2C_0F_2 + b_3C_0F_3 + b_4C_0F_4 + b_5C_1F_1 + b_6C_1F_2 + b_7C_1F_3 + b_8C_1F_4 \qquad (7)$$

$$\begin{array}{cccccccc} C_0F_1 & C_0F_2 & C_0F_3 & C_0F_4 & C_1F_1 & C_1F_2 & C_1F_3 & C_1F_4 \\ \begin{bmatrix}1\\0\\0\\0\\0\\0\\0\\0\end{bmatrix} & \begin{bmatrix}0\\1\\0\\0\\0\\0\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\1\\0\\0\\0\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\1\\0\\0\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\1\\0\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\0\\1\\0\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\0\\0\\1\\0\end{bmatrix} & \begin{bmatrix}0\\0\\0\\0\\0\\0\\0\\1\end{bmatrix} \end{array}$$

Table 6.6 shows the results for the all-differences model and the all-means model. The all-differences model includes seven interaction variables and excludes the variable for not-college-graduate parent/two biological parents. The unit vector is included instead of this variable. As a result, the intercept is the mean for that category.

The all-means model includes the dummy variable for not-college-graduate parent/two biological parents in place of the unit vector and, thus, estimates all means. The dummy variable coefficients in Model 1 can be obtained by subtracting the means in Model 2 from mean for the not-college-graduate-parent/two-biological-parent group.

Table 6.6 Interactions Between Family Structure and College-Graduate Parent With Math Score as Dependent Variable

	Model	
	1	2
Independent Variable	***B***	***B***
Intercept	49.27*	—
Not College Grad. × Two Bio.	—	49.27
Not College Grad. × Bio./Step.	−1.09*	48.18
Not College Grad. × Single	−1.08*	48.19
Not College Grad. × Other Fam.	−2.32*	46.95
College Grad. × Two Bio.	7.00*	56.27
College Grad. × Bio./Step.	4.44*	53.71
College Grad. × Single	5.17*	54.44
College Grad. × Other Fam.	4.55*	53.82
R^2	.125	—

*p < .05.

Interactions Between Dummy Variables and an Interval Variable

Modeling interactions between a set of dummy independent variables and an interval-level independent variable is simpler than the case of interactions between two sets of dummy variables because there is only one within-group effects model to consider.[4]

4. This book does not discuss interactions between two interval variables. I find that these interactions are difficult to interpret. As an alternative, I suggest dividing into categories the interval variable that it is conditioned on and then using the procedure for interactions between dummy variables and an interval variable. For a brief discussion of interactions between interval variables, see Gordon (2010), and for more extensive treatments, see Jaccard (1990, 2001).

Equation 8 is the standard interaction model for the interaction between family structure and socioeconomic status (SES). F_1 is two biological parents, F_2 is biological parent/stepparent, F_3 is single parent, and F_4 is other family. SES is parental SES quartile and has four values:

$$Y = a + b_1F_2 + b_2F_3 + b_3F_4 + b_4\text{SES} + b_5F_2\text{SES} + b_6F_3\text{SES} + b_7F_4\text{SES} \tag{8}$$

U	F_2	F_3	F_4	SES	F_2SES	F_3SES	F_4SES
1	0	0	0	1	0	0	0
1	0	0	0	2	0	0	0
1	0	0	0	3	0	0	0
1	0	0	0	4	0	0	0
1	1	0	0	1	1	0	0
1	1	0	0	2	2	0	0
1	1	0	0	3	3	0	0
1	1	0	0	4	4	0	0
1	0	1	0	1	0	1	0
1	0	1	0	2	0	2	0
1	0	1	0	3	0	3	0
1	0	1	0	4	0	4	0
1	0	0	1	1	0	0	1
1	0	0	1	2	0	0	2
1	0	0	1	3	0	0	3
1	0	0	1	4	0	0	4

In Equation 8, the interactions are formed by multiplying each dummy variable times the interval variable. Examination of the data matrices that follow Equation 8 shows that the interaction variables are all nested within the SES variable. This means the coefficients for the interaction variables will estimate differences from the coefficient for SES. The coefficient for the SES variable will capture the SES effect for the subgroup not covered by the interactions, those in two-biological-parent families:

$$Y = a + b_1F_2 + b_2F_3 + b_3F_4 + b_4F_1\text{SES} + b_5F_2\text{SES} + b_6F_3\text{SES} + b_7F_4\text{SES} \tag{9}$$

Equation 9 for the within-group effects model uses F_1SES rather than SES as in the standard interaction model. Examination of the following data matrices shows that the within-group effects model basically takes the interval SES variable and breaks it into four subparts, one for each family structure group:

U	F_2	F_3	F_4	F_1SES	F_2SES	F_3SES	F_4SES
1	0	0	0	1	0	0	0
1	0	0	0	2	0	0	0
1	0	0	0	3	0	0	0
1	0	0	0	4	0	0	0
1	1	0	0	0	1	0	0
1	1	0	0	0	2	0	0
1	1	0	0	0	3	0	0
1	1	0	0	0	4	0	0
1	0	1	0	0	0	1	0
1	0	1	0	0	0	2	0
1	0	1	0	0	0	3	0
1	0	1	0	0	0	4	0
1	0	0	1	0	0	0	1
1	0	0	1	0	0	0	2
1	0	0	1	0	0	0	3
1	0	0	1	0	0	0	4

Model 3 in Table 6.7 is the within-group effects interaction model, and the SES coefficient is largest for two biological parents and smaller for the other three groups.

Table 6.7 Interactions Between Family Structure and Parental SES With Math Score as Dependent Variable

	Model		
	1	2	3
Independent Variable	***B***	***B***	***B***
Bio./Step.	−1.58*	.04	.04
Single	−.53*	.64	.64
Other Fam.	−1.59*	−1.24*	−1.24*
SES QUAR	3.32*	3.52*	—
Two Bio. × SES QUAR	—	—	3.52*
Bio./Step. × SES QUAR	—	−.64*	2.88*
Single × SES QUAR	—	−.51*	3.01*
Other Fam. × SES QUAR	—	−.11	3.41*
Intercept	43.75	43.23	43.23
R^2	.147	.147	.147

*p < .05.

Model 2 is the standard interaction model and shows that the SES coefficient for other family is not significantly different than the coefficient for two biological parents, whereas the coefficients for the other two groups are significantly smaller than the coefficient for two biological parents.

It is possible to estimate a third-order difference in a two-way interaction model. A first-order difference is a simple difference. A second-order difference is a difference in a difference. Thus, a third-order difference is a difference in a difference in a difference. In Equations 10 and 11, the INTSUM variable is used to estimate third-order differences. INTSUM sums up the three interaction variables:

$$Y = a + b_1F_2 + b_2F_3 + b_3F_4 + b_4\text{SES} + b_5\text{INTSUM} + b_6F_3\text{SES} + b_7F_4\text{SES} \tag{10}$$

$$\text{INTSUM} = F_2\text{SES} + F_3\text{SES} + F_4\text{SES} \tag{11}$$

U	F_2	F_3	F_4	SES	INTSUM	F_3SES	F_4SES
1	0	0	0	1	0	0	0
1	0	0	0	2	0	0	0
1	0	0	0	3	0	0	0
1	0	0	0	4	0	0	0
1	1	0	0	1	1	0	0
1	1	0	0	2	2	0	0
1	1	0	0	3	3	0	0
1	1	0	0	4	4	0	0
1	0	1	0	1	1	1	0
1	0	1	0	2	2	2	0
1	0	1	0	3	3	3	0
1	0	1	0	4	4	4	0
1	0	0	1	1	1	0	1
1	0	0	1	2	2	0	2
1	0	0	1	3	3	0	3
1	0	0	1	4	4	0	4

I replace F_2SES in Equation 9 with INTSUM in Equation 10. The coefficients for F_2SES, F_3SES, and F_4SES in Equation 9 estimated interaction effects. However, the F_3SES and F_4SES variables in Equations 10 and 11 are nested in the INTSUM variable. Thus, each coefficient now estimates the difference in the interaction effect for that variable from the interaction effect for F_2SES.

Model 1 in Table 6.8 is a standard interaction model and estimates the interaction of SES quartile with family structure. The three family structure variables are multiplied times SES to calculate the interaction variables. In Model 2, the Bio./Step. × SES variable is replaced by the INTSUM variable, which is the sum of the three interaction variables in Model 1.

Table 6.8 Interactions Between Family Structure and Parental SES With Math Score as Dependent Variable

	Model	
	1	2
Independent Variable	***B***	***B***
Bio./Step.	.03	.03
Single	.64*	.64*
Other Fam.	−1.24*	−1.24*
SES QUARTILE	3.52*	3.52*
Bio./Step. × SES QUAR.	−.64*	—
INTSUM	—	−.64*
Single × SES QUAR.	−.51*	−.13
Other Fam. × SES QUAR.	−.11	.53
Intercept	43.23	43.23
R^2	.147	.147

*p < .05.

The coefficient for the INTSUM variable in Model 2 of Table 6.8 is the same as the coefficient for Bio./Step. × SES in Model 1. When we replace a standard interaction variable such as Bio./Step. × SES with a "summer" variable like INTSUM, the summer variable will have the same coefficient as the variable that it replaced. The coefficients that change are the variables that previously were not nested in the Bio./Step. × SES variable but are now nested in the INTSUM variable. The coefficients for Single × SES and Other × SES, which estimated interaction effects like the coefficient for Bio./Step. × SES in Model 1, now estimate differences from the interaction effect for Bio./Step. × SES in Model 2.

Although the within-group effects model and the standard interaction model are basic tools for analyzing interactions, using a variable like INTSUM is likely to be a rare occurrence for most researchers. However, the example using the INTSUM variable illustrates how a "summer" variable can estimate a third-order difference, a difference of a difference of a difference. Use of the INTSUM variable also illustrates a general property of interaction variables. That is, although we can create only one set of interaction variables with two independent variables, the number of models we can create with this set of interaction variables is varied. When someone says, "Run the interactions," the appropriate response should be "Which model?"

Three-Way Interactions

In regression modeling, researchers typically use control models, but they also often use two-way interaction models. Three-way interactions models are rare in published research. The many additional variables in a three-way interaction models make interpretation difficult. Where would a researcher even start in interpreting such a model? Model 1 in Table 6.9 has four dummy variables and is an additive model. Model 2 adds seven interaction variables and is a three-way interaction model and seems complex. Model 2 has a mixture of variables, and it is not immediately obvious how to analyze such a complex model.

Table 6.9 Interactions Among Race/Ethnicity, Female, and College-Graduate Parent With Math Score as Dependent Variable

	Model	
	1	2
Independent Variable	*B*	*B*
Black	−4.54*	−3.68*
Other Race/Ethnicity	.16	.06
College-Graduate Parent	6.80*	7.17*
Female	−.13	.25
College Grad. × Female	—	−.84*
Black × College Grad.	—	−3.68*
Other × College Grad.	—	.71
Black × Female	—	−.08
Other × Female	—	−.44
Black × College Grad. × Fem.	—	2.87*
Other × College Grad. × Fem.	—	.25
Intercept	49.13	48.96
R^2	.136	.138

*$p < .05$.

The challenge of interpreting a three-way interaction model is illustrated in the models in Table 6.9. The additive model includes two dummy variables for race/ethnicity, a single dummy variable for college-graduate parent, and a single dummy variable for female. The three way-interaction model then adds seven interaction variables to the additive model. There is one two-way interaction variable for the interaction between college-graduate parent and female. There are two two-way interaction variables for the interaction between race/ethnicity and college-graduate parent. There are two two-way interactions for the interaction between race/ethnicity and female. Finally, there are two interaction variables for the three-way interaction among race/ethnicity, college-graduate parent, and female. The three-way interaction model more than doubles the number of variables in the additive model. Interpreting this model seems difficult as a result of the number and variety of variables.

In two-way interaction models, the standard interaction model estimates differences between coefficients in the within-group effects model. Thus, it is necessary to choose the within-group effects model of interest to interpret the standard interaction model properly. We can take a similar approach in working with a three-way interaction model.[5] There will also be a standard interaction model and a within-group effects model in the three-way interaction case.

Specifying the within-group model for three-way interactions involves focusing on one set of two-way interactions. Although there are three alternative two-way models in this example, the focus in Model 2 in Table 6.10 is on the two-way interaction between college-graduate parent and female. Model 3 is the same as the three-way interaction model in Table 6.9, but the variables are listed in a specific order. The three-way interaction model can be viewed as taking the two additive variables and the two-way interaction for college-graduate parent and female and then interacting those three variables with race/ethnicity.

Model 1 in Table 6.11 is the standard interaction model. At the top of the column are the variables for Black and other, which will serve as the conditioning variables. Next are variables for college-graduate parent, female, and the two-way interaction between those variables. Again, the focus is on the two-way interaction between college-graduate parent and female when interpreting the three-way interaction analysis. Model 1 also includes six additional variables that capture the three-way interaction.

Within Model 2 in Table 6.11 are standard interaction models for the interaction between college-graduate parent and female for Whites, Blacks, and other, separately. Thus, Model 1 can be viewed as testing whether the interactions between college-graduate parent and female are different for Blacks and other as compared with Whites. Model 2, on the other hand, shows the two-way interaction model for each race/ethnic group.

---

5. Jaccard (2001), pages 24–30, suggests that one approach to making three-way interactions more interpretable is to view three-way interactions as two-way interactions conditioned on a third variable.

Table 6.10 Interactions Among Race/Ethnicity, Female, and College-Graduate Parent With Math Score as Dependent Variable

	Model		
	1	2	3
Independent Variable	***B***	***B***	***B***
Black	−4.54*	−4.54*	−3.68*
Other Race/Ethnicity	.16	.17	.06
College-Graduate Parent	6.80*	7.04*	7.17*
Female	−.13	.08	.25
College Grad. × Female	—	−.48	−.84*
Black × College Grad.	—	—	−3.68*
Black × Female	—	—	−.08
Black × College Grad. × Fem.	—	—	2.87*
Other × College Grad.	—	—	.71
Other × Female	—	—	−.44
Other × College Grad. × Fem.	—	—	.25
Intercept	49.13	49.03	48.96
R^2	.136	.136	.138

*$p < .05$.

The key to a three-way interaction analysis is whether the three-way interaction coefficient is significantly different from zero. In this particular analysis, the key variable is the Black × college-graduate parent × female interaction. The three-way interaction involving other race/ethnicity is not substantively interesting.

The coefficient for the Black × college-graduate parent × female variable is significant, and interpretation of the three-way interaction model is appropriate. Examining the within-group two-way interactions in Model 2 in Table 6.11 helps greatly in interpreting the three-way interaction. The three-way interaction coefficient in Model 1 is 2.87. This is the difference between the coefficient two-way interaction between college-graduate parent and female for Blacks and the coefficient for Whites (2.03 – (–.84)). These two-way interaction coefficients can be interpreted in two ways. I interpret them as the additional female effect for those with a college-graduate parent.

Table 6.11 Interactions Among Race/Ethnicity, Female, and College-Graduate Parent With Math Score as Dependent Variable

	Model				
	1	2	3	4	5
Independent Variable	***B***	***B***	***B***	***B***	***B***
Black	−3.68*	−3.68*	—	—	—
Other Race/Ethnicity	.06	.06	—	—	—
College-Graduate Parent	7.17*	—	7.17*	3.49*	7.88*
Female	.25	—	.25	.17	−.19
College Grad. × Female	−.84*	—	−.84*	2.03*	−.60
White × College Grad.	—	7.17*	—	—	—
White × Female	—	.25	—	—	—
White × College Grad. × Fem.	—	−.84*	—	—	—
Black × College Grad.	−3.68*	3.49*	—	—	—
Black × Female	−.08	.17	—	—	—
Black × College Grad. × Fem.	2.87*	2.03*	—	—	—
Other × College Grad.	.71	7.88*	—	—	—
Other × Female	−.44	−.19	—	—	—
Other × College Grad. × Fem.	.24	−.60	—	—	—
Intercept	48.96	48.96	48.96	45.28	49.02
R^2	.138	.138	.117	.063	.126

*p < .05.

Also, the last three columns in Table 6.11 show models for the two-way interaction between college-graduate parent and female for Whites only in Model 3, for Blacks only in Model 4, and for other race/ethnicity only in Model 5. The coefficients in these three models are the same as the coefficients in the within-group model, Model 2. Thus, the standard interaction model, Model 1, tests whether the respective coefficients in Models 3, 4, and 5 are equal. This means that running separate interaction models is equivalent to running interaction models with the race/ethnicity variable as the conditioning variable.

Table 6.12 Interactions Among Race/Ethnicity, Female, and College-Graduate Parent With Math Score as Dependent Variable

	Model		
	1	2	3
Independent Variable	***B***	***B***	***B***
Black	−3.68*	−3.68*	−3.68*
Other Race/Ethnicity	.06	.06	.06
College-Graduate Parent	7.17*	—	—
Female	.25	—	—
College Grad. × Female	−.84*	—	—
White × College Grad.	—	7.17*	7.17*
White × Female	—	.25	—
White × Not College Grad. × Fem.	—	—	.25
White × College Grad. × Fem.	—	−.84*	−.59*
Black × College Grad.	−3.68*	3.49*	3.49*
Black × Female	−.08	.17	—
Black × Not College Grad. × Fem.	—	—	.17
Black × College Grad. × Fem.	2.87*	2.03*	2.20*
Other × College Grad.	.71	7.88*	7.88*
Other × Female	−.44	−.19	—
Other × Not College Grad. × Fem.	—	—	−.19
Other × College Grad. × Fem.	.24	−.60	−.79*
Intercept	48.96	48.96	48.96
R^2	.138	.138	.138

*$p < .05$.

In the discussion on two-way interactions, I suggested that the best way to interpret a two-way interaction is to interpret the within-groups model. Table 6.12 shows a further developed within-groups model for the three-way interaction. In this model, there is a separate female effect for those not with a college-graduate parent and for those with a college-graduate parent within each race/ethnic group. The female effect for Whites not with a college graduate parent is not significant, whereas the female effect for Whites with a college-graduate parent is significantly negative. In contrast, the female effect for Blacks not with a college-graduate parent is not significant, whereas the female effect for Blacks with a college-graduate parent is significantly positive.

White females with a college-graduate parent score lower in math than males do, but the reverse is true for Blacks. Black females with a college-graduate parent score higher in math than do males. A college-graduate parent disadvantages females among Whites but advantages females among Blacks. In both groups, there is no sex difference for those not with a college-graduate parent.

Estimating Separate Models

In my analysis of three-way interactions, I showed that we can view a three-way interaction model as a set of two-way interaction models conditioned on a third variable. We can use this modeling approach to test whether coefficients for subgroups in separate models are equal. For example, a researcher may use regression modeling to arrive at a specific model and then look to see whether the coefficients in the model are different for Whites, Blacks, and other race/ethnicity.[6] One approach to determining whether the models are different from one another is the Chow test.[7] However, we can conduct an equivalent test that uses interactions.

Table 6.13 shows separate models for Whites, Blacks, and other race ethnicity for the effect of family structure, college-graduate parent, and family income on math scores.

The results indicate that although there is a significant positive effect of two-biological-parent family for White, there is no significant effect for Blacks. In addition, the effect for college-graduate parent is less for Blacks than Whites, but the effect of income is larger for Blacks than for Whites.

The method of using interactions to estimate separate models interacts all the independent variables with the conditioning variable.[8] In this case, race/ethnicity is the conditioning variable and we interact each of the three dummy variables representing race/ethnicity with the other independent variables.

Table 6.14 shows the results of regression involving interacting race/ethnicity with the other independent variables. Model 1 is the additive model, and Model 2 shows the

---

6. Hardy (1993), page 49, and Jaccard (2001), page 17, make the important point that the problem with estimating separate models and comparing coefficients across models is that no significance test for the difference between coefficients is conducted.

7. Demaris (2004), pages 110–112, and Gordon (2010), pages 277–286, present the Chow test for testing the difference between separate models. I feel using the *F* test for linear regression or the chi-square test for logistic regression for testing the difference in fit between the additive model and the with-in group effects model is a simpler approach.

8. Demaris (2004), pages 151–152, and Hill, Griffiths, and Judge (1997), pages 190–192, discuss how to use interactions to do a global test for the difference between similar regression models for two or more groups.

Table 6.13 Separate Models for Whites, Blacks, and Other Race/Ethnicity for Effects of Family Structure, College-Graduate Parent, and Family Income With Math Score as Dependent Variable

	White	Black	Other
Independent Variable	***B***	***B***	***B***
Two-Biological-Parent Family	1.15*	.81	1.39*
College-Graduate Parent	5.48*	3.31*	6.15*
Family Income	.14*	.17*	.17*
Intercept	47.61	44.30	47.36
R^2	.141	.080	.151

*p < .05.

standard interaction model that allows the effects of two-biological-parent family, college-graduate parent, and family income to vary by race/ethnicity. When we focus on the White/Black contrast, we observe that Model 2 indicates that the difference in the effects of two-biological-parent family and family income are not significantly different from one another. However, the coefficient for the Black × college-graduate parent is significant, indicating that the effect of college-graduate parent is less for Blacks than for Whites.

Model 3 in Table 6.14 is the within-group effects interaction model. Notice that the coefficients for two-biological-parent family, college-graduate family, and family income for Whites, Blacks, and other race/ethnicity are the same as those for the separate models in Table 6.13. Model 4 is what I call the "separate models model." This model is estimated by suppressing the intercept and adding a dummy variable for White to the model. The coefficients for two-biological-parent family, college-graduate parent, and family income are the same as in Model 3. The difference between Model 3 and Model 4 is that the coefficients for Black and other race/ethnicity in Model 4 are no longer additive variables as in Model 3 but are now the same as the intercepts for the separate models as in Table 6.12.

I do not recommend estimating the separate models model except for the purpose of learning and exploring regression modeling. However, the separate models model does show that it is possible to estimate one regression model that replicates exactly the results from running separate models.

Example Using Logistic Regression

Family income has an important influence on whether a student attends a private high school. Those with families with higher incomes are more likely to attend since private high schools often cost substantially more than public schools.

Table 6.14 Interactions Among Race/Ethnicity and Family Structure, College-Graduate Parent, and Family Income With Math Score as Dependent Variable

	Model			
	1	2	3	4
Independent Variable	***B***	***B***	***B***	***B***
White	—	—	—	47.61
Black	−3.98*	−3.31*	−3.31*	44.30
Other Race/Ethnicity	.41*	−.26	−.26	47.36
Two-Biological-Parent Family	1.19*	1.15*	—	—
College-Graduate Parent	5.49*	5.48*	—	—
Family Income	.15*	.14*	—	—
White × Two-Parent Family	—	—	1.15*	1.15*
White × College-Graduate Parent	—	—	5.48*	5.48*
White × Family Income	—	—	.14*	.14*
Black × Two-Parent Family	—	−.34	.81	.81
Black × College-Graduate Parent	—	−2.17*	3.31*	3.31*
Black × Family Income	—	.03	.17*	.17*
Other × Two-Parent Family	—	.24	1.39*	1.39*
Other × College-Graduate Parent	—	.67*	6.15*	6.15*
Other × Family Income	—	.03*	.17*	.17*
Intercept	47.46	47.61	47.61	—
R^2	.160	.162	.162	—

*$p < .05$.

The analysis in Table 6.15 considers whether the effect of family income ($10,000 units) is the same for Blacks as for Whites. Model 1 shows a significant effect of income in the additive model. Model 2 is the standard interaction model and shows a significantly higher effect for family income for Blacks than for Whites. The coefficient for family income in Model 3 is .08 for Blacks and .06 for Whites. Each additional unit of family income has a greater impact on Black chances than on those of Whites.

Table 6.15 Interactions Among Race/Ethnicity and Family Income With Private High School as Dependent Variable

	Model		
	1	2	3
Independent Variable	*B*	*B*	*B*
Black	−.13	−.30*	−.30*
Other Race/Ethnicity	−.21*	−.22*	−.22*
Family Income	.06*	.06*	—
White × Fam. Inc.	—	—	.06*
Black × Fam. Inc.	—	.02*	.08*
Other × Fam. Inc.	—	.00	.06*
Intercept	−2.15	−2.14	−2.14
−2 log-likelihood	16,639.7	16,632.0	16,632.0

*$p < .05$.

The analysis in Table 6.16 considers the interaction between race/ethnicity and whether the parent is a college graduate on chances of attending private high school. When the variables involved in the interaction are both categorical, the researcher has a choice about which variable to condition on the other. In the analysis in Table 6.16, college-graduate parent is conditioned on race/ethnicity.

The coefficient for Black × College Grad. is significant and negative. This indicates that the effect of college-graduate parent is less for Blacks than for Whites. Model 3 in Table 6.16 shows that the effect of college-graduate parent is 1.46 for Whites and 1.18 for Blacks. Having a college-graduate parent is less of an advantage for Blacks than for Whites.

In Table 6.17, race/ethnicity is conditioned on college-graduate parent. The coefficient for Black × College Grad. in Model 2 is significant and negative. This indicates that the Black coefficient is more negative among those who have a college-graduate parent than among those who do not have a college-graduate parent. Model 3 shows that the Black coefficient among those who do not have a college-graduate parent is −.03 and not significant and that the coefficient for those with a college-graduate parent is −.31 and significant. There is a racial difference in chances of attending private high school among those who have a college-graduate parent but not among those who do not have a college-graduate parent.

Table 6.16 Interactions Between Race/Ethnicity and College-Graduate Parent With Private High School as Dependent Variable

	Model		
	1	2	3
Independent Variable	*B*	*B*	*B*
Black	–.19*	–.03	–.03
Other Race/Ethnicity	–.22*	–.15*	–.15*
College-Graduate Parent	1.40*	1.46*	—
White × College Grad.	—	—	1.46*
Black × College Grad.	—	–.28*	1.18*
Other × College Grad.	—	–.11	1.35*
Intercept	–2.30	–2.34	–2.34
–2 log-likelihood	16,757.1	16,752.4	16,752.4

*$p < .05$.

Table 6.17 Interactions Between Race/Ethnicity and College-Graduate Parent With Private High School as Dependent Variable

	Model		
	1	2	3
Independent Variable	*B*	*B*	*B*
College-Graduate Parent	1.40*	1.46*	1.46*
Black	–.19*	–.03	—
Other Race/Ethnicity	–.22*	–.15*	—
Not College Grad. × Black	—	—	–.03
Not College Grad. × Other	—	—	–.15*
College Grad. × Black	—	–.28*	–.31*
College Grad. × Other	—	–.11	–.26*
Intercept	–2.30	–2.34	–2.34
–2 log-likelihood	16,757.0	16,752.4	16,752.4

*$p < .05$.

Summary

An additive model constrains the effect of an independent variable to be the same in categories of a second independent variable. Estimating interaction models allows the researcher to remove this constraint and to allow the effects of an independent variable to vary for subgroups.

I argue that for any standard interaction model, there is a within-group effects model that serves as the basis for the standard interaction model. The within-group effects interaction model estimates the effect of an independent variable within subgroups defined by a second independent variable. The standard interaction model then provides the test for whether the effect of an independent variable is equal across subgroups. I suggest that a common reason that the standard interaction model can be misinterpreted or can be difficult to interpret is a result of the lack of a clearly defined and recognized within-group effects interaction model.

This chapter does not address interactions between two interval variables. I find that interactions between two interval variables are difficult to interpret. Since the coefficient is not readily interpretable, a fuller understanding of the meaning of the interaction coefficient for two interval variables is best achieved by graphing the relationship measured by the interaction coefficient.[9] In contrast, I show in this chapter that interactions involving independent variables represented by dummy variables are readily interpretable. Therefore, I suggest that a more interpretable way to analyze interactions between two interval variables is to convert one of the interval variables into a categorical variable captured by dummy variables. I would suggest at least three or four categories for the recoded interval variable to model the effects of the variable adequately. Although such an approach is less parsimonious because a variable whose effect on the dependent variable was captured with one coefficient is now captured with multiple coefficients, the advantage of interpretability outweighs, in my view, any loss of parsimony.

When doing regression analysis using interaction variables, the researcher must pay close attention to the sample size in each sample subgroup. One rule of thumb is that if a sample size of 25 is needed to get a good estimate of a population mean, then any subgroups as defined by an interaction variable need to have sample sizes of at least 25. So after creating a dummy variable that represents an interaction, look at the frequency distribution for the variable to make sure that both of the categories defined by the dummy variable have a sample size of at least 25.

Key Concepts

conditional differences: the idea that the difference between two groups on a dependent variable will vary within categories of a third variable.

two-way interaction: a variable constructed by multiplying two independent variables; when one of the variables involved is a dummy variable, then the interaction variable will be a subset of one or both of the independent variables involved in the interaction.

9. Gordon (2010) provides a brief treatment of interactions between interval variables. Jaccard (1990, 2001) provide more extensive treatments.

standard interaction model: an interaction model that incorporates interactions between two sets of independent variables; the interaction coefficients in a standard interaction model measure the additional effect of one of the independent variables for one group compared with the effect of the same independent variable for another group when at least one of the sets of independent variables is measured with dummy variables.

within-group effects interaction model: an interaction model that incorporates interactions between two sets of independent variables; the interaction coefficients in a within-group effects interaction model measure the effect of one of the independent variables for one group and the effect of that same independent variable for another group when at least one of the sets of independent variables is measured with dummy variables.

three-way interactions: a variable constructed by multiplying three independent variables; to enhance interpretability, I suggest that researchers view a three-way interaction as a set of two-way interactions conditioned on a third independent variable.

separate models interaction model: a two-way interaction model that estimates in one model separate regression models for subgroups defined by a third independent variable; the objective of the discussion of the separate models model was to show that interactions can be used to estimate the same coefficients that can be obtained by estimating separate regressions for subgroups.

Chapter Exercises

1. Replicate the regressions and the table for the interaction between race/ethnicity and college-graduate parent with private as the dependent variable like in Table 6.14. Use PRIVATE, BLACK, OTHRACE, PARCOLL, WPARCOLL, BPARCOLL, and OPARCOLL in the analysis.
2. Run regressions using the linear regression procedure, and create a table similar to the one for the interaction between race/ethnicity and college-graduate parent with math score as the dependent variable. Use X2TXMTSCOR, BLACK, OTHRACE, PARCOLL, WPARCOLL, BPARCOLL, and OPARCOLL in the analysis.

   How do the effects of college-graduate parent compare for Whites and Blacks?
3. Run regressions using the linear regression procedure, and create a table for the interaction between family structure and family income with math score as the dependent variable like in Table 6.7. You will use the following variables in creating the models: two biological parent, biological parent/stepparent, single, other family, and family income. Use X2TXMTSCOR, STEP, SINGLE, FAMOTH, FAMINC, TWOINC, STEPINC, SINGLEINC, and FAMOTHINC in the analysis.

   How do the effects of family income differ for those in two-biological-parent, biological-parent/stepparent, and single-parent families?