CHAPTER 2

BETWEEN-PERSON ANALYSIS AND INTERPRETATION OF INTERACTIONS

Chapter 2 presents some of the necessary building blocks that we'll need to have in place before diving into longitudinal analysis. It uses a single continuous example to meet two goals. The first goal is to review general linear models for **between-person analysis**, in which each person has only one outcome (i.e., cross-sectional data) and thus only one model residual (i.e., the difference between the observed and model-predicted outcome). Second, this chapter will describe how to specify and interpret interactions among continuous or categorical predictors in general linear models. Although interactions are a default in some variants of general linear models (e.g., ANOVA), they are less common in others (e.g., regression), particularly among continuous predictors. Because interactions play a prominent role in the longitudinal models in the rest of this text, though, it is important to understand them thoroughly before proceeding any further. Thus, this chapter tackles interaction effects in familiar general linear models for cross-sectional data prior to presenting them within more complex longitudinal models.

1. Between-Person (Cross-Sectional; Between-Groups) Analysis

1.A. Decomposing General Linear Models

In the phrase **general linear model**, the term *general* means that we are assuming all model residuals have a normal distribution with a mean of 0 and some estimated variance that is constant over persons and across the values of any predictors. Critically, the model residuals are assumed to be independent of each other as well, which is why longitudinal data will require a different model. The term *linear* means that a linear combination of the predictors is used to create an expected outcome, in which the contribution of each predictor is weighted by an estimated slope parameter that describes the size of its unique relationship with the outcome.

Finally, the term *model* implies that we are not just trying to *describe* the outcome, but that we are trying to *predict* the outcome using other information (i.e., the model predictors). This chapter is intended to serve only as a refresher of general linear models, with the goal of reinforcing the concepts and vocabulary that will be necessary for us to move forward into more complex models. Before continuing, though, it is necessary to discuss the different names possessed by the different general linear model variants and how they relate to each other.

Traditionally, general linear models including continuous predictors are called **regression models,** in which the slope of the line relating each continuous predictor to a continuous outcome is estimated. Of interest in regression models is the size and direction of each of these slopes, how these slopes differ after including slopes for additional predictors, and how much variance in the outcome is explained by the predictors. In contrast, general linear models including categorical (grouping) predictors are called **analyses of variance models** (or **ANOVA models**). Of interest in ANOVA models is the size and direction of mean differences between groups as well as how much variation in the outcome can be attributed to those mean differences. General linear models with both continuous and categorical predictors are called **analysis of covariance models** (or **ANCOVA models**) if the continuous predictors are allowed only main effects, or *regression* models if the continuous predictors are allowed to interact with other predictors.

Ultimately, though, these naming conventions serve only to maintain the arbitrary distinctions among variants of what is essentially just one kind of model—a *between-person general linear model,* as defined above. Many of the distinctions among these models arise from the use of different statistical routines (e.g., SAS PROC REG vs. GLM, or SPSS Regression vs. GLM). These routines differ by default in how they represent the effects of categorical predictors and in how they summarize and evaluate model effects in their output.

In addition, regression and ANOVA models have historically been taught using different mathematical representations for the sake of convenience, and this may be another reason why their underlying communalities are not always readily apparent. For instance, because predictors in regression models are treated as continuous variables, they have many possible values that can create many different expected outcomes. As a result, regression models are usually summarized using equations that show how the combination of model predictors, each weighted by its slope, creates an expected outcome for each observation (i.e., a predicted outcome for each person). This can be a very general and useful way of expressing a model. In contrast, ANOVA models are not commonly expressed using equations that predict individual outcomes. This may be because the same expected outcome would be received by all individuals within the same group (or crossing of groups in designs with multiple grouping variables), and so simply describing the model via differences in group means is a more direct way to convey ANOVA results. Thus, although ANOVA models could be expressed using individual prediction equations like in regression, it can be less convenient to do so. However, because the models we will cover in the rest of this text will include continuous and categorical predictors, for continuity the models will be

presented as equations that describe how a predicted outcome is created for each observation.

Furthermore, because ANOVA models rely on least squares estimation, their presentation usually emphasizes how evaluation of group mean differences operates through sums of squares, mean squares, *F*-ratios, and other statistics that are fundamental to least squares estimation. And because ANOVA models are special cases of regression models, these summary statistics also appear in evaluating the overall quality of a regression model (such as an *F*-test for whether the variance accounted for by a set of predictors is significantly different than 0). But because the models in the rest of this text will require other methods of estimation (such as *maximum likelihood,* as described in subsequent chapters), we will emphasize interpretation of the fixed effects for the predictors and residual variance components *per se,* rather than the least squares routes by which they are obtained. In cross-sectional analyses with complete data, least squares and maximum likelihood estimation will result in the same model estimates anyway.

Finally, there is another salient difference in how the effects of predictors are typically specified in regression versus ANOVA models. In regression models, predictors are included primarily as main effects, such that their slope is assumed to be linear and constant over all other predictors. Although interactions can be included among predictors, this is not the default model specification. In contrast, whenever two or more grouping variables are included in ANOVA, all possible interactions among them tend to be estimated by default. As a result, it has become standard practice in ANOVA to interpret both main effects and interaction effects, and much attention is given to decomposing interaction effects via contrasts of specific group means. This difference in typical procedure leads to a few important distinctions between ANOVA and regression in the way main effects and interactions effects are most often interpreted.

Specifically, in anticipation of including interaction effects in ANOVA, the main effects of grouping variables are usually coded with **contrasts,** or by coding the predictor variable such that the mean across all possible predictor groups is 0. (These concepts will be illustrated in more detail later in the chapter.) For instance, given equal group sizes, a two-group variable may be represented as –0.5 for one group and +0.5 for the other group, so that the mean across groups is 0. This way, if the main effects and interaction effects of multiple grouping variables are included in an ANOVA model, the main effects can be interpreted as the *overall mean difference* across the levels of each grouping variable (i.e., averaged across all other predictors), or what is known as a **marginal main effect.** For groups with more than two levels (and especially when different sizes of the groups are to be represented), this coding can be tedious, but software packages will take care of all such coding so as to retain marginal main effects in the presence of interactions.

In contrast, regression software requires the user to specify how the predictor effects are to be entered into the model. For instance, given two continuous predictors, we can multiply them together and include their product to represent an interaction between them in the model (as demonstrated later in the chapter). Unlike ANOVA models in which the grouping variables are coded such that the

mean across groups is 0, the mean of the continuous predictors is usually not 0 by default. As a result, although interactions can be interpreted the same as in ANOVA, the main effects cannot. Rather than remaining marginal main effects (such that the main effect is interpreted as the average effect across all other predictors), the main effects become **simple main effects**, such that they are interpreted as *the main effect specifically when the interacting predictor is 0*. These distinctions can be confusing, and so the examples in this chapter are designed to help illustrate them more fully. Suffice to say for now that in general, when learning regression models, the interpretation of interactions (and the impact they have on the main effects of the predictors) is not as heavily emphasized as when learning ANOVA models, and as a result, the consequences of including interactions between continuous predictors in regression models may be less well understood. But because many of the fixed effects to be interpreted in longitudinal models will be interactions, the goal of this chapter is to remedy any misconceptions in interpreting main effects and interactions in between-person (cross-sectional) models first.

To summarize, this chapter will review underlying concepts and vocabulary across general linear model variants in order to build a common language with which to move forward. Persons wanting a more thorough discussion of general linear models *per se* should consult any of the excellent texts that address these models, such as Maxwell and Delaney (2004) for ANOVA, and Cohen, Cohen, West, and Aiken (2002) for regression. We now illustrate between-person general linear models that include main effects of continuous and categorical predictors.

1.B. A Between-Person Empty Model

As introduced in chapter 1, all statistical models have two sides: The model for the means and the model for the variance. The **model for the means** (i.e., fixed effects, structural model) describes how the predictors are weighted and combined to create an expected outcome for each observation. The **model for the variance** (i.e., random effects and residuals, stochastic model) describes how the model residuals (the difference between the actual outcome for each observation and the outcome predicted by the model for the means) are distributed and related to each other. All general models, no matter how complex they become, begin with one term for each side of the model, as in the **between-person empty model** shown in Equation (2.1):

$$y_i = \beta_0 + e_i \tag{2.1}$$

in which y_i is the outcome for individual *i*. The model for the means contains just a single fixed effect: an intercept, β_0 (pronounced "beta zero"). An **intercept** is defined as the expected outcome when all predictors = 0. But because there are no predictors in Equation (2.1)—hence the name *empty model*—the intercept β_0 is just the grand mean of y_i. That is, if we know nothing else about a person, our best naïve guess for his or her y_i outcome is its grand mean. This is the clearest example of where the term *model for the means* comes from—in the empty model in Equation (2.1), the model for the means literally contains just the grand mean, β_0.

When we fit this model to an outcome variable, we receive an estimate of what β_0 is for that sample along with its **standard error** (to be described shortly) that indicates how precise the intercept estimate is. The model for the variance contains just a single **residual** (i.e., error term), e_i, which is the difference between the predicted and the actual outcome for individual *i*. Here we note a critical distinction between the model for the means and the model for the variance regarding their focus of interest. Unlike the fixed effects (in the model for the means) that each receive an estimate with a standard error, in the model for the variance, each term (just e_i here) is used to represent a *variance* that is estimated instead. That is, rather than estimating an e_i residual value for each person, the model focuses on providing an estimate of the *variance of the e_i values* across the sample of *N* persons given *k* fixed effects, given by Equation (2.2):

$$\sigma_e^2 = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N - k} \tag{2.2}$$

in which the variance of the e_i values, typically denoted as σ_e^2 and called **residual variance** or **error variance**, represents all the unknown reasons why the observed y_i outcome differs from the predicted y_i outcome (known as $\hat{y}_i$, pronounced "y hat") for each person. The $y_i - \hat{y}_i$ deviations are then squared, summed over *N* persons, and that quantity is divided by *N* persons minus the *k* number of fixed effects (including the fixed intercept). In a between-person general linear model, we assume that the e_i residuals are normally distributed with a mean = 0 and a variance = σ_e^2, and that the e_i residuals are also independent with constant variance across persons and predictors. The models for longitudinal data in the rest of the text will modify these assumptions of independence and constant variance, but not the assumption about normality of the e_i residuals (although chapter 13 will have more to say about this topic).

The empty model in Equation (2.1) represents the starting point of every statistical model that could possibly follow. From the perspective of predicting the outcome, it is absolutely the worst we can do—all the variance in the y_i outcome is yet to be accounted for because everyone is predicted to have its grand mean (that is, $\hat{y}_i$ is predicted from only β_0 so far, or *k* fixed effects = 1). Thus, the next logical step is to include predictors that might help create more accurate predicted outcomes and reduce or explain the σ_e^2 residual variance.

To illustrate the empty model and those that follow, consider the following example: A researcher is interested in describing individual differences in cognitive functioning. To suit the goals of the chapter, data were generated for a single occasion based loosely on patterns found in the *Octogenarian Twin Study of Aging* (OCTO, a longitudinal study described in chapter 1). Our example data include 550 older adults age 80 to 97 years (*M* = 84.93, *SD* = 3.43). Cognition was assessed by the *Information Test,* a measure of general world knowledge (i.e., crystallized intelligence; *M* = 24.82, *SD* = 10.99, range = 0 to 44). For simplicity, the Information Test outcome will be called *cognition* throughout this example. Although we could use software for general linear models (e.g., SAS or SPSS GLM, STATA REGRESS), we will instead use restricted maximum likelihood estimation within the more flexible software for general linear

Table 2.1 Results from between-person models including main effects only. Bold values are $p < .05$.

Model Parameters		*Equation 2.3: Empty Model*			*Equation 2.4: Add Age*			*Equation 2.6: Add Grip Strength*			*Equation 2.7: Add Sex*			*Equation 2.8: Add Dementia*		
		Est	*SE*	*p <*	*Est*	*SE*	*p <*	*Est*	*SE*	*p <*	*Est*	*SE*	*p <*	*Est*	*SE*	*p <*
Model for the Means																
β_0	Intercept	**24.82**	0.47	.001	**24.78**	0.46	.001	**24.70**	0.45	.001	**26.96**	0.74	.001	**29.26**	0.70	.001
β_1	Age (0 = 85 years)				**–0.55**	0.13	.001	**–0.42**	0.13	.002	**–0.43**	0.13	.001	–0.41	0.12	.001
β_2	Grip Strength (0 = 9 lbs)							**0.80**	0.15	.001	**0.55**	0.17	.001	**0.60**	0.15	.001
β_3	Sex (0 = Men, 1 = Women)										**–3.80**	0.99	.001	**–3.66**	0.89	.001
	Dementia Group															
β_4	None vs. Future													**–5.72**	1.02	.001
β_5	None vs. Current													**–16.48**	1.52	.001
$\beta_5 - \beta_4$	Future vs. Current													**–10.76**	1.71	.001
Model for the Variance																
σ_e^2	Residual Variance	120.76			117.46			112.12			109.38			88.07		
	R^2 relative to Empty Model				.03			.07			.10			.27		

mixed models to be featured in the rest of the text (e.g., SAS, SPSS, or STATA MIXED), of which the general linear models are a special case. Table 2.1 shows the results for each incremental model in this section.

Although atypical when conducting a regression analysis, we will begin by estimating the empty between-person model in Equation (2.1) for our example data, as shown in Equation (2.3):

$$\text{Cognition}_i = \beta_0 + e_i \tag{2.3}$$

in which Cognition_i is the Information Test outcome for individual *i*. As shown in the first set of columns in Table 2.1, the estimate we obtain for the intercept $\beta_0 = 24.82$ (with a standard error, or SE = 0.47) exactly matches the grand mean of Cognition_i in this sample. We also obtain an estimate of the variability of the e_i residuals of $\sigma^2_e = 120.76$ (typically labeled as *Mean Square Error* or MSE in GLM output). The square root of this variance exactly matches the standard deviation reported for the cognition outcome as well. Not surprisingly, the variance accounted for in cognition (given as R^2 in the output) is exactly 0, because σ^2_e still contains all possible variance in cognition (i.e., $\hat{y}_i$ is based only on the grand mean β_0 so far). Although uninformative in a predictive sense, this empty model does provide a useful baseline for further models, in that by knowing how much outcome variance there is in the first place, we can more directly see how the predictors we subsequently include will reduce this variation. This will be especially helpful in later chapters.

1.C. Between-Persons Analysis Using Continuous Predictors

To continue our example, many characteristics can potentially relate to cognition, but perhaps a reasonable place to start in this sample of older adults is chronological age. Thus, we can expand the model shown in Equation (2.3) to include a predictor for age (*M* = 84.93 years, *SD* = 3.43, range = 80 to 97 years), as shown in Equation (2.4):

$$\text{Cognition}_i = \beta_0 + \beta_1(\text{Age}_i - 85) + e_i \tag{2.4}$$

in which the model for the means now contains two fixed effects (β_0 and β_1). The model for the variance still contains just e_i, which still represents the difference between the observed outcome and the outcome predicted by the model for the means for each person. In this model, though, e_i is the discrepancy in the actual cognition outcome that remains after predicting cognition from age, now given by $\beta_0 + \beta_1(\text{Age}_i - 85)$ instead of just β_0. As seen in the second set of columns in Table 2.1, the variance of the e_i residuals was reduced to $\sigma^2_e = 117.46$. The R^2 value for reduction in σ^2_e that would be reported from GLM output is calculated as the model sum of squares divided by the total sums of squares from the model plus error. After adding a main effect of age, this is a reduction of approximately 3% relative to the empty model. Another way to arrive at approximately this same figure that will generalize to later models is by calculating the proportion reduction in σ^2_e relative to that of the empty model, or $R^2 = (120.76 - 117.46)/120.76 = .03$ here.

The new fixed effect β_1 is a **slope**, defined as the difference in the expected outcome for a one-unit difference in the predictor. The reason that $Age_i - 85$ was included as a predictor rather than Age_i was to keep the intercept interpretable. That is, because the intercept is the expected outcome when all predictors are 0, the scale of each predictor should include a meaningful 0 point. In this case, because our example includes data from adults age 80 to 97 years, an intercept at $Age_i = 0$ would fall far outside the range of the data (i.e., it would be the expected cognition outcome at birth). Accordingly, we changed the scale of our age predictor so that it includes 0 by **centering:** We subtracted a constant from each person's age so that 0 would fall within the range of the new age predictor. Thus, 85 years is the new 0 for age in the model. Age 85 was chosen for the centering point because it is near the sample mean, but other ages observed in the sample (e.g., age 80 or 90) could have been chosen as well. Given the scaling of age in years, $\beta_1 = -0.55$ is the expected difference in cognition for a one-unit difference in age: for each year older, cognition is expected to be lower by 0.55. The age slope is assumed to be linear, such that a one-unit difference in age has the exact same effect on cognition at all ages. Although nonlinear effects of age could also be added (e.g., age^2 to represent a quadratic effect of age), the models in this example will include linear slopes for continuous predictors only.

Whether or not the age slope $\beta_1 = -0.55$ is significantly different from 0 depends on its **standard error**, which can be derived using the formula in Equation (2.5):

$$SE_{\beta_x} = \sqrt{\frac{Var(y_i) * (1 - R_Y^2)}{Var(x_i) * (1 - R_X^2) * (N - k)}} = \sqrt{\frac{120.76 * (1 - 0.03)}{11.75 * (1 - 0) * (550 - 2)}} = 0.13 \qquad (2.5)$$

in which the β_1 SE = 0.13, as calculated using the original variance in the cognition outcome (120.76), the proportion reduction in cognition variance from the model for the means (0.03), and the variance in the age predictor (11.75). More generally, Equation (2.5) shows how the standard error of any fixed effect (SE of β_x) depends on a few key pieces of information. First, in the numerator, the SE depends on how well the model for the means can predict the outcome, as indexed by the amount of outcome variance that remains. All things being equal, fixed effects in models for the means that account for more outcome variance will have smaller SE values. Second, the denominator serves to scale that remaining outcome variance based on the scale of the original x_i predictor for which we are deriving an SE. More specifically, the denominator starts with the original amount of variance in the x_i predictor, multiplied by how much of its variance can be accounted for by the other predictors in the model (i.e., the reciprocal of its VIF, variance inflation factor). So far, because age is the only predictor, its R^2 from other predictors is 0, and so its total variance is then multiplied by the *N* sample size minus *k* fixed effects (*k* = 2 for β_0 and β_1). Thus, given the same amount of remaining outcome variance, to the extent that a predictor has more variance in general, or has less shared variance with the other predictors in the model, the SE for its effect will be smaller. Finally, by taking the square root, the entire quantity is transformed from a variance metric to a

standard deviation metric (i.e., SE is the *standard deviation* of the sampling distribution for the fixed effect, rather than its *variance*).

In general for any fixed effect, the ratio of its estimate divided by its SE is distributed as a *t*-statistic, or what is known as a **Wald test.** Here, $t = -0.55/0.13 = -4.23$ (within rounding error). That *t*-statistic can be compared to a *t*-distribution to determine the probability (the *p*-value) that the β_1 age slope estimate is different from 0. The degrees of freedom for the *t*-statistic here is $N - 2$ for the two fixed effects (β_0 and β_1). Thus, relative to a critical value of ±1.96 for $p < .05$, the age slope estimate β_1 is significantly different than 0. Said differently, the 3% of the variance in cognition that age accounted for was a statistically significant reduction.

In addition to evaluating whether a given fixed effect is significantly different than 0, we can also form a confidence interval around any fixed effect using its standard error. For instance, a *95% confidence interval* around the age slope β_1 can be found as (β_1 ± 1.96*SE). That is, if the study was replicated numerous times, the confidence interval would include the true value of the age slope 95% of the time. Thus, smaller SE values lead to narrower intervals, or less expected variability in the size of the effect across samples. In this example, the confidence interval for the age slope would be: CI = –0.55 ± 1.96*0.13, or –0.80 to –0.30, within rounding error. The fact that the interval does not overlap 0 also means the age slope is significantly different from 0.

The other fixed effect, the intercept, was estimated as $\beta_0 = 24.78$ (SE = 0.46), which can be interpreted as the expected cognition outcome for someone who is 85 years old (i.e., given that the intercept is the expected outcome when all predictors are 0, which is 85 for $\text{Age}_i - 85$). Had we not centered Age_i at 85, the intercept would have been estimated as $\beta_0 = 71.20$ (SE = 11.46), which would have been the expected cognition outcome at birth, an impossible (and thus highly imprecise) value of cognition given its range of 0 to 44. However, the age slope β_1 and the error variance σ_e^2 would still be the same, because *centering* (i.e., subtracting a constant from the predictor variable) does not change the predictor's main effect, so long as no interactions with the predictor are included in the model. We will elaborate on this point in the sections to come.

In studies with older adults, age is usually included as a control variable prior to examining the effects of other predictors. But another factor that may relate to cognition is physical condition—more frail individuals may have diminished cognition. We can examine this idea in our example data by including a predictor of *grip strength,* measured in pounds per square inch ($M = 9.11$ pounds, $SD = 2.99$, range = 0 to 19 pounds). After centering grip strength at 9 pounds, we can then add the centered predictor to the model, as shown in Equation (2.6):

$$\text{Cognition}_i = \beta_0 + \beta_1(\text{Age}_i - 85) + \beta_2(\text{Grip}_i - 9) + e_i \tag{2.6}$$

the results for which are shown in the third set of columns in Table 2.1. The model for the means now contains three fixed effects: β_0, β_1, and β_2. The intercept $\beta_0 = 24.70$ (SE = 0.45) is the expected cognition for someone who is both age 85 and has a grip strength of 9 pounds (i.e., when $\text{Age}_i - 85$ and $\text{Grip}_i - 9$ are both 0). The age slope

is still the expected difference in cognition for a one-unit difference in age, but is now $\beta_1 = -0.42$ (SE = 0.13): for each additional year of age, cognition is expected to be significantly lower by 0.42. The age slope is closer to 0 than in the previous age-only model in Equation (2.4) because it now reflects the *unique contribution of age* holding constant the grip strength of each person. Because age and grip strength are correlated in the sample ($r = -.18$), the contribution of age after controlling for grip strength is smaller (with a slightly larger SE due to a reduction in the amount of unique age variance unrelated to grip strength). The grip strength slope $\beta_2 = 0.80$ (SE = 0.15) indicates that for each additional pound of grip strength (holding age constant), cognition is expected to be significantly higher by 0.80. Finally, the variance in cognition that remains after controlling for age and grip strength is $\sigma^2_e = 112.12$, which has now been reduced by 7% relative to the original variance in cognition as given by the empty model, $R^2 = (120.76 - 112.12) / 120.76 = .07$.

1.D. Between-Person Analysis Using Categorical Predictors

So far we have only considered continuous predictors of cognition (i.e., a *regression*). Now we consider categorical or grouping predictors as well (i.e., *analysis of covariance*). Continuing with our example, another factor that might relate to cognition is sex—there may be differences in cognition between men and women. Our example data is 41.27% men and 58.73% women. Because there are only two groups, we can represent the difference between them with a single variable. To make sure the intercept stays interpretable, we include a dummy-coded predictor for sex such that 0 = men and 1 = women, as shown in Equation (2.7):

$$\text{Cognition}_i = \beta_0 + \beta_1(\text{Age}_i - 85) + \beta_2(\text{Grip}_i - 9) + \beta_3(\text{SexMW}_i) + e_i \tag{2.7}$$

results for which are shown in the fourth set of columns in Table 2.1. The estimated fixed effect for the sex difference of $\beta_3 = -3.80$ (SE = 0.99) indicates that women ($\text{SexMW}_i = 1$) are predicted to have significantly lower cognition by 3.80 than men ($\text{SexMW}_i = 0$). As a result of including the sex difference β_3, the intercept $\beta_0 = 26.96$ (SE = 0.74), which is now the expected cognition specifically for a man who is age 85 and has 9 pounds of grip strength. The residual variance has been reduced to $\sigma^2_e = 109.38$, or a total reduction of 9% relative to the empty model, $R^2 = (120.76 - 109.38)/120.76 = .09$. The age slope β_1 and the grip strength slope β_2 are still significant after controlling for differences between men and women in cognition. However, the unique effect of grip strength is reduced ($\beta_2 = 0.80$ vs. $\beta_2 = 0.55$) with a higher SE due to the correlation between grip strength and sex (i.e., there are sex differences in grip strength favoring men, $r = -.40$), whereas the unique effect of age and its SE are similar with or without sex in the model given the low correlation between age and sex ($r = .05$).

The choice to represent the sex predictor such that men were the reference (0) group was arbitrary; other versions of the sex predictor could also have been used. For instance, if we had dummy-coded the sex predictor such that women were the

reference group instead (i.e., 0 = Women, 1 = Men), then the intercept would be $\beta_0 = 26.96 - 3.80 = 23.16$. Another alternative is *effects coding,* in which 0 becomes the mean of the grouping variable (i.e., as more commonly used in ANOVA). In our example, had we coded men as –0.4127 and women as 0.5873 to match the proportion of men and women in the sample, then the intercept would have been $\beta_0 = 25.39$ instead, which could then be interpreted as the intercept averaged across men and women (but also still conditional on age = 85 and grip strength = 9 pounds). If we wished to give each group equal weight instead of weighting based on sample size, group values of ±0.5 could have been used instead. Regardless of the coding of the sex predictor, however, so long as there is a difference of exactly 1.0 between the two possible values of the sex predictor, the main effect of sex representing the difference between men and women and the rest of the model estimates will stay the same, because centering does not change the model-predicted outcome.

In addition to age, grip strength, and sex, the final predictor of cognition we will consider is a dementia diagnosis during the rest of the longitudinal study. Specifically, we will evaluate differences among three types of persons: those who will *not* be diagnosed with dementia (*none* group = 1; 72.55%), those who will *eventually* be diagnosed with dementia later in the study (*future* group = 2; 19.82%), and those who *already* have been diagnosed with dementia (*current* group = 3; 7.64%). Although there are three possible differences among the three groups, only two group differences need to be represented in the model, as the third is redundant (i.e., it could be determined by the other two group differences). Given that it is the largest in the sample, we will select the *none* group as our reference by creating two new variables to represent the difference between the *none* group and the other groups: $DemNF_i$ (none = 0, future = 1, current = 0) and $DemNC_i$ (none = 0, future = 0, current = 1). We then include both dementia group contrasts as predictor variables simultaneously in the model, as shown in Equation (2.8):

$$\begin{aligned}\text{Cognition}_i = {} & \beta_0 + \beta_1(\text{Age}_i - 85) + \beta_2(\text{Grip}_i - 9) + \beta_3(\text{SexMW}_i) \\ & + \beta_4(\text{DemNF}_i) + \beta_5(\text{DemNC}_i) + e_i\end{aligned} \tag{2.8}$$

in which Cognition_i is still the outcome for individual *i,* as now predicted by age, grip strength, sex, and the two grouping variables for dementia diagnosis. Results from the model in Equation (2.8) are shown in the fifth set of columns in Table 2.1; each effect was significant.

The *none* group is the reference because it is the only group that has a 0 for both DemNF_i and DemNC_i. Thus, the intercept $\beta_0 = 29.26$ is now the expected cognition outcome for a man who is age 85, who has 9 pounds of grip strength, and who will not be diagnosed with dementia. By including both group contrasts simultaneously as predictors, we can interpret them as the difference between the reference group and the alternative group coded 1 for each contrast. Thus, the slope for DemNF_i $\beta_4 = -5.72$ indicates that relative to persons who will not be diagnosed with dementia (none), those who will be diagnosed with dementia (future) are expected to have significantly lower cognition by 5.72. Likewise, the slope for DemNC_i $\beta_5 = -16.48$ indicates that relative to persons who will not be diagnosed with dementia (none),

those who have already been diagnosed with dementia (current) are expected to have significantly lower cognition by 16.48. The slopes for the other predictors remained significant after controlling for dementia diagnosis. Although the other predictor variables are related to dementia group, the SE values for their slopes are still smaller because of the reduction in the residual variance after including effects of dementia group. Specifically, the residual variance has been reduced to $\sigma_e^2 = 88.07$, or a total reduction of 27% relative to the empty model, $R^2 = (120.76 - 88.07) / 120.76 = .27$.

The manual coding of group differences in dementia diagnosis using $DemNF_i$ and $DemNC_i$ is not typically how such categorical grouping variables are specified in a general linear model (i.e., as in ANOVA). When differences between groups are coded manually as we've done here, the model reports significance tests for each specific group contrast separately. But what may be of interest instead is whether there are significant differences across the groups in general—this omnibus effect is what a typical ANOVA would report instead of (or in addition to) the separate group contrasts. To obtain this omnibus information for the model in Equation (2.8), we would remove the two dementia group contrasts we created, and instead indicate that the original three-category dementia variable is a categorical predictor within the program syntax (i.e., on the CLASS statement in SAS, on the BY statement in SPSS, or using the i. option in STATA). After doing so, we obtain an omnibus overall test (i.e., a multivariate Wald test with two degrees of freedom) of whether there is a significant difference across the three groups, $F(2, 544) = 67.06$, $p < .001$.

This designation of a categorical grouping variable is also convenient in that any desired comparisons between groups can then be requested (not just those that are explicitly given by the manual contrast variables in the model). For instance, our model has only given us two of the three possible group differences—we do not know yet if the future and current dementia groups also differ significantly. One way to obtain this contrast is to make future dementia the reference group by replacing $DemNF_i$ and $DemNC_i$ with new contrasts of $DemFN_i$ (none = 1, future = 0, current = 0) and $DemFC_i$ (none = 0, future = 0, current = 1) and re-estimating the model. But this is not necessary if your software provides estimates and standard errors for any fixed effect that is *implied* by the model, even if not given directly by a model parameter. These statements (e.g., ESTIMATE in SAS, TEST in SPSS, LINCOM in STATA, or NEW in M*plus*, as included in the syntax online) are much more convenient than changing the reference group and re-estimating the model. Using this approach here, we can obtain the model-implied difference between the future and current groups as $\beta_5 - \beta_4 = -16.48 + 5.72 = -10.76$ (SE = 1.71, $p < .001$).

However, an unfortunate side effect is that the group differences provided directly within the model may not be what you had intended—in SAS and SPSS, they are relative to the group coded *highest* numerically or last alphabetically; in STATA, they are relative to the group coded *lowest* numerically (although this can be changed). As such, you should be extra cautious in assessing differences between groups when the program is in charge of creating the contrasts instead of you! These issues are further elaborated in the appendix at the end of this chapter.

2. Interpreting Interactions Among Continuous Predictors

So far we have assumed that the effects of age, grip strength, sex, and dementia group are additive. But what if age differences in cognition are greater in those with worse grip strength? Similarly, what if the sex difference in cognition favoring men that we found earlier depends on dementia group? The general idea that the effect of a model predictor "depends on" another model predictor is referred to more generally as **moderation**, which is tested by including an interaction term between the predictors whose effects are thought to depend on one another (see Aiken & West, 1991). Interaction variables may need to be created in advance depending on the particular software routine used. This is yet another reason why we are using general linear mixed modeling procedures in our statistical packages, which generally do not require that interaction terms be created ahead of time—instead, interaction effects can be estimated directly in the syntax by specifying a special character (such as an asterisk in SAS or SPSS or a hashtag in STATA) between the predictor variables that will interact in the model.

Interpretation of interactions and their constituent main effects has historically been fraught with difficulty, primarily resulting from confusion as to how those main effects should then be interpreted. Some authors have suggested that main effects should *not* be interpreted when they are included in an interaction, but I will take a decidedly different perspective—*main effects can and should be interpreted, especially when included in an interaction*. The trick is to interpret the main effects correctly! As described in the next sections, the correct way to interpret main effects is *conditionally on their interacting predictor,* not marginally, as when they are included only as main effects. Using our working example predicting cognition, we next illustrate how to interpret interactions among continuous predictors, followed by interactions among categorical predictors, and then interactions among continuous and categorical predictors. For clarity, new interaction effects are underlined in each of the model equations that follow.

To provide an example of moderation between continuous predictors, we add to our previous model an interaction effect between age and grip strength, as shown in Equation (2.9):

$$\begin{aligned}\text{Cognition}_i = {} & \beta_0 + \beta_1(\text{Age}_i - 85) + \beta_2(\text{Grip}_i - 9) + \beta_3(\text{SexMW}_i) \\ & + \beta_4(\text{DemNF}_i) + \beta_5(\text{DemNC}_i) + \underline{\beta_6(\text{Age}_i - 85)(\text{Grip}_i - 9)} + e_i \end{aligned} \quad (2.9)$$

in which an additional slope β_6 has been added to represent the interaction of age and grip strength. Fixed effects from the model in Equation (2.9) are shown in the first set of columns in Table 2.2. Some of the effects originally present in the main effects model in Equation (2.8) now take on different interpretations due to the age by grip strength interaction. We will discuss each of these effects in turn. Throughout this section, small differences (i.e., ≤ 0.01) between the values calculated in the text and those reported in the tables may occur due to rounding error.

To begin, some of the fixed effects in the model for the means are interpreted the same as in the previous main effects only model in Equation (2.8). These include

Table 2.2 Fixed effects from models with an age by grip strength interaction. Bold values are $p < .05$.

Model Parameters		*Equation 2.9 when Age = 85, Grip = 9*			*Equation 2.9 when Age = 80, Grip = 12*			*Equation 2.9 when Age = 90, Grip = 6*		
		Est	*SE*	*p <*	*Est*	*SE*	*p <*	*Est*	*SE*	*p <*
β_0	Intercept	**29.41**	0.69	.001	**31.09**	1.09	.001	**24.03**	1.15	.001
β_1	Age									
	Grip Strength (0 = 6 lbs)							**–0.70**	0.15	.001
	Grip Strength (0 = 9 lbs)	**–0.33**	0.12	.006						
	Grip Strength (0 = 12 lbs)				0.04	0.19	.851			
β_2	Grip Strength									
	Age (0 = 80 years)				0.00	0.25	.986			
	Age (0 = 85 years)	**0.62**	0.15	.001						
	Age (0 = 90 years)							**1.23**	0.26	.001
β_3	Sex (0 = Men, 1 = Women)	**–3.46**	0.89	.001	**–3.46**	0.89	.001	**–3.46**	0.89	.001
	Dementia Group									
β_4	None vs. Future	**–5.92**	1.01	.001	**–5.92**	1.01	.001	**–5.92**	1.01	.001
β_5	None vs. Current	**–16.30**	1.51	.001	**–16.30**	1.51	.001	**–16.30**	1.51	.001
$\beta_5 - \beta_4$	Future vs. Current	**–10.38**	1.70	.001	**–10.38**	1.70	.001	**–10.38**	1.70	.001
β_6	Age by Grip Interaction									
	Age (0 = 85) by Grip (0 = 9)	**0.12**	0.04	.003						
	Age (0 = 80) by Grip (0 = 12)				**0.12**	0.04	.003			
	Age (0 = 90) by Grip (0 = 6)							**0.12**	0.04	.003

the intercept β_0, the main effect of sex β_3, and the main effects of dementia diagnosis group β_4 and β_5. Although these fixed effects are now the unique effects after controlling for the age by grip strength interaction, their interpretations do not change because they are not part of the interaction. Because no new predictors have been added to the model, the intercept β_0 is still the expected cognition outcome for an 85-year-old man with 9 pounds of grip strength who will not be diagnosed with dementia. Likewise, because they are not part of an interaction, the main effects of sex β_3 and dementia group β_4 and β_5 continue to represent their group mean differences in cognition. We would use the terms **unconditional main effect** or **marginal main effect** to describe the main effects of sex and dementia group, in that their effects do not depend on the value of any other predictor because they are not part of an interaction with any other predictors—not yet, anyway!

What have changed in the model for the means after adding the age by grip strength interaction β_6, however, are the interpretations of the main effects for age β_1 and grip strength β_2. Previously, these slopes indicated the expected difference in cognition for a one-unit difference in age or grip strength, respectively, and these

effects were expected to hold equally across the sample (i.e., they were *unconditional* when they were not included in an interaction). As seen by comparing the last set of columns in Table 2.1 to the first set of columns in Table 2.2, after adding an interaction between them, the main effect of age β_1 changed from –0.41 to –0.33 and the main effect of grip strength β_2 changed from 0.60 to 0.62. Although small in this instance, these changes are necessary and expected because main effects that are part of an interaction can no longer be considered main effects. Instead, they become **simple (conditional) main effects,** such they become the effect of the predictor *specifically when their interacting predictor = 0.* Here, because of the age by grip strength interaction, the simple main effect of age β_1 is now the *age slope when grip strength = 0.* Because centered grip strength = 0 is really 9 pounds, the age slope β_1 indicates that *specifically for someone with 9 pounds of grip strength,* for every year older, cognition is expected to be significantly lower by 0.33. Likewise, the simple main effect of grip strength β_2 is now the *grip strength slope when age = 0.* Because centered age = 0 is really 85 years, the grip strength slope β_2 indicates that *specifically for someone who is age 85,* for every additional pound of grip strength, cognition is expected to be significantly higher by 0.62.

2.A. Implications of Centering for Interpreting Simple Main Effects of Interactions

In general, the correct way to interpret main effects that are included in an interaction is to make a *conditional* (not marginal) interpretation—they become *simple* main effects that apply only when the interacting predictor is 0. For this reason, it is imperative that the predictors have a meaningful 0 value. Otherwise, the intercept and main effects can be nonsense. For instance, in this example, what if we had not centered our predictors—what if we had used the original values of age and grip strength and their interaction instead? The intercept β_0 and the simple main effects of age β_1 and grip strength β_2 would change radically, because they would then be conditional on the 0 values of the original predictors instead. Thus, the intercept β_0 would be the expected cognition outcome at birth (age = 0 years) for someone with absolutely no grip strength (grip strength = 0 pounds). Likewise, the simple age main effect β_1 would describe age differences in cognition specifically for persons with absolutely no grip strength (grip strength = 0 pounds), and the grip strength simple main effect β_2 would describe grip strength differences in cognition specifically at birth (age = 0 years). By centering age at 85 years and grip strength at 9 pounds, we ensure that the intercept (and the simple main effects of their interaction) are evaluated conditionally for persons who actually exist in our data, and therefore are useful to us. The most typical centering point for a predictor is its grand mean, so that the intercept and main effects are evaluated where there is the most data (i.e., the center of the predictor's distribution). In reality, however, any centering constant within the observed scale of the predictor can be used to facilitate interpretation of the intercept and the simple main effects of an interaction, and sometimes the grand mean may not be as useful as other constants with more inherent meaning.

Consider, for example, a predictor of years of education when measured in adults. In a sample of persons with high school degrees, college degrees, or graduate degrees, the grand mean for years of education could be something like 14.67 years. Although you could use 14.67 years of education as a reference point, it's an odd choice because no one in the sample is likely to have provided 14.67 years as a response. In this case, it may be more meaningful to pick 12 years of education as a centering point instead, such that the reference group would become persons who graduated high school, or perhaps 16 years of education, such that the reference group would become persons who graduated college. Such absolute centering points can also be helpful for emphasizing the absolute values of predictors within a given sample, as well as why potentially different findings may be reported for a predictor whose absolute values differ across samples. For instance, "mean education" could be 10 years in one sample but 18 years in another sample, and the simple main effect of an interacting predictor may look very different when evaluated as conditional on 10 years of education rather than conditional on 18 years of education. Interpreting interacting main effects as conditional simply on "mean education" can mask such important distinctions when the mean education differs across samples; centering at an absolutely meaningful value (like 12 years of education) can emphasize them instead.

There is nothing inherently wrong with using any centering point of your choosing. You can subtract from each predictor the grand mean, a meaningful constant, or perform no centering at all, and the model will still account for the same amount of outcome variance and predict the same expected outcomes. That is, your results will not become incorrect as a consequence of centering or not centering, but your coefficients will be strange if the 0 values of your predictors extend beyond the possible range of their scales. Simply put, the problem with not centering is that the intercept and simple main effects of an interaction may not provide useful information. For this reason, I strongly recommend centering each predictor so that the intercept is an interpretable and meaningful value as evaluated when the predictor is 0, but this is especially important when evaluating simple main effects that are conditional on their interacting predictor.

2.B. Interaction Coefficients Modify Their Simple Main Effects

Now that we have considered how the main effects become conditional on the interacting predictor, it's time to interpret the significant interaction coefficient itself. As shown in the first set of columns in Table 2.2, the interaction coefficient from the model in Equation (2.9) was $\beta_6 = 0.12$. Its inclusion reduced the residual variance from $\sigma_e^2 = 88.07$ to 86.76 (for a total reduction of approximately 28% of the original variation in cognition, $R^2 = (120.76 - 86.76)/120.76 = .28$.

The trick to correctly interpreting interaction coefficients is to remember their role in the model. More specifically, whereas the role of main effects is to adjust the *intercept* (i.e., a positive main effect makes the expected outcome or intercept go up, a negative main effect makes the intercept go down), the *role of two-way interactions*

is to adjust the slopes of the simple main effects. That is, the interaction effect operates only indirectly on the intercept by adjusting the simple main effects (which then adjust the intercept more directly). Given two interacting simple main effects, there are two possible ways to interpret the interaction, which would describe how each simple main effect depends on the value of the other predictor. Both interpretations will always be correct, but one may be more convenient to present than the other. That is, there is no way to distinguish statistically which is "the moderator"—that is a distinction to be made in interpretation, because the interaction moderates both main effects at once.

Accordingly, one way to interpret the age by grip strength interaction β_6 is to describe how the *effect of age depends on grip strength*. To do so, we start with the simple main effect of age $\beta_1 = -0.33$ as evaluated specifically when grip strength = 9 pounds (because 9 is the new 0 in the centered grip strength predictor). For every additional pound of grip strength, *the age slope β_1 becomes less negative by the interaction of $\beta_6 = 0.12$*. Thus, the interaction *weakens* the effect of age, such that the β_1 slope for the expected change in cognition for each year of age is *less negative* in stronger people (smaller by 0.12 per pound of grip strength). To put it in more simply in plain English, we would say that age matters *less* for predicting cognition in stronger people.

The other way to interpret the age by grip strength interaction is to describe how the effect of grip strength depends on age. To do so, we start with the simple main effect of grip strength $\beta_2 = 0.62$ as evaluated specifically when age = 85 years (because age 85 is the new 0 in the centered age predictor). For every additional year of age, *the grip strength slope β_2 becomes more positive by the interaction of $\beta_6 = 0.12$*. Thus, the interaction also *strengthens* the effect of grip strength, such that the β_2 slope for the expected change in cognition for each pound of grip strength is *more positive* in older people (larger by 0.12 per year of age). Or, to put it in English, we would say that strength matters *more* for predicting cognition in older people.

This example also illustrates an important point about interaction coefficients—we cannot simply look at the direction of the interaction to determine its influence in the model. Here, the positive interaction coefficient $\beta_6 = 0.12$ served to make the age slope $\beta_1 = -0.33$ *less negative* (weaker by 0.12 per additional pound of grip strength), Furthermore, because *less negative* is in the same direction as *more positive,* the positive interaction effect implies that not only would the negative age slope become less negative as grip strength increases, but that the age slope would eventually become positive in very strong people. At the same time, however, the positive interaction $\beta_6 = 0.12$ made the grip strength slope $\beta_2 = 0.62$ *more positive* (stronger by 0.12 per additional year of age), but it also implies that the grip strength slope would become *less positive* as age decreases, eventually becoming negative in much younger people. In contrast, had the interaction been *negative,* it could have made the negative main effect of age more negative (stronger), or it could have made the positive main effect of grip strength less positive (weaker). Thus, interaction effects must be interpreted relative to their simple main effects in the model.

2.C. Re-Centering Main Effects to Decompose Interactions

We can see more directly how the model predicts what the simple main effects of age and grip strength will be for any value of the interacting predictor by using a little bit of calculus. More specifically, if we take the first derivative of the function given in Equation (2.9) with respect to age and then with respect to grip strength, the result is shown in Equation (2.10):

$$
\begin{aligned}
&\text{Age Slope} = \beta_1 + \beta_6(\text{Grip}_i - 9)\\
&\text{Age Slope at Grip Strength} = 6: \quad \beta_{1new} = -0.33 + (0.12 * -3) = -0.70\\
&\text{Age Slope at Grip Strength} = 9: \quad \beta_{1new} = -0.33 + (0.12 * 0) = -0.33\\
&\text{Age Slope at Grip Strength} = 12: \quad \beta_{1new} = -0.33 + (0.12 * 3) = -0.04
\end{aligned}
\tag{2.10}
$$

$$
\begin{aligned}
&\text{Grip Strength Slope} = \beta_2 + \beta_6(\text{Age}_i - 85)\\
&\text{Grip Strength Slope at Age} = 80: \quad \beta_{2new} = 0.62 + (0.12 * -5) = 0.00\\
&\text{Grip Strength Slope at Age} = 85: \quad \beta_{2new} = 0.62 + (0.12 * 0) = 0.62\\
&\text{Grip Strength Slope at Age} = 90: \quad \beta_{2new} = 0.62 + (0.12 * 5) = \ 1.23
\end{aligned}
$$

in which the simple main effects of age under different centering points for grip strength (6, 9, or 12 pounds) as well as the simple main effects of grip strength under different centering points for age (80, 85, or 90 years) have been calculated. For each additional pound of grip strength, the age slope β_1 becomes less negative by the interaction $\beta_6 = 0.12$, and for each additional year of age, the grip strength slope β_2 becomes more positive by the interaction $\beta_6 = 0.12$.

Fortunately, we can also calculate any model-implied simple main effect without using calculus, and it involves three steps. First, we would examine the model equation and extract any terms that include the predictor whose simple slope we wish to find. For instance, to find the simple slope for age, we would examine Equation (2.9), and extract only the following terms that include the age predictor: $\beta_1(\text{Age}_i - 85) + \beta_6(\text{Age}_i - 85)(\text{Grip}_i - 9)$. Second, we would factor out the predictor variable—here, this would result in: $(\text{Age}_i - 85)[\beta_1 + \beta_6(\text{Grip}_i - 9)]$. Third, the term that then multiplies the predictor in [ ] then becomes its new slope—as shown in the first line of Equation (2.10). This logic and process generalizes to models with higher-order interactions (i.e., three-way and four-way interactions) as well.

Although we can use these slope equations to calculate what the simple main effects would be under any centering constant, we cannot easily obtain their SEs (and their significance tests) this way. But we can do so simply by re-estimating the model using different centering points for each interacting predictor. For instance, in the first set of columns in Table 2.2, age is centered at 85 years and grip strength is centered at 9 pounds. Given the negative correlation between age and grip strength, I opted to pair younger and stronger persons to create a second reference point, and to pair older and weaker persons to create a third reference point. Thus, in the second set of columns in Table 2.2, age is centered at 80 years and grip strength is centered at

12 pounds. In the third set of columns in Table 2.2, age is centered at 90 years and grip strength is centered at 6 pounds. The additional age values of 80 and 90 years were chosen as specific values of interest within the ages sampled, and the grip strength values of 6 and 12 pounds were chosen as ±1 standard deviation (SD) around the mean of 9 pounds. By re-estimating the same model using different centering points for the simple main effects and interaction, we can obtain estimates and SEs for the simple main effects *as evaluated at specific values of the interacting predictors*. Although it is common to use ±1 SD of the predictors as points with which to re-center, any meaningful value can be used to decompose an interaction in this fashion.

Let us examine the results in Table 2.2. The interaction of age and grip strength $\beta_6 = 0.12$ is the same across models with differing centering points because it is the highest order effect—it does not depend on any other predictor. Likewise, the main effects of sex β_3 and of dementia group β_4 and β_5 are the same as well because they are not part of the interaction. Although not shown, the amount of outcome variance accounted for is the same across models with different centering points, because centering will not change the predictions of a model.

But what have changed are the intercept β_0, simple main effect of age β_1, and simple main effect of grip strength β_2 because they are conditional on the 0 value of each predictor. The fixed intercept varied from 24.03 to 29.41 to 31.09 across models because it is the expected cognition outcome when age = 0 and grip strength = 0, and the location of 0 depends on which centering point was used (e.g., 80, 85, or 90 years of age; 6, 9, or 12 pounds for grip strength). In each case, however, the intercept is also still specifically for a man who will not be diagnosed with dementia given that the reference groups for sex and dementia are the same. Similarly, the age slope β_1 varied from –0.70 (for 6 pounds of grip strength, third set of columns) to –0.33 (for 9 pounds of grip strength, first set of columns) to 0.04 (for 12 pounds of grip strength, second set of columns). Likewise, the grip strength slope β_2 varied from 0.00 (for age 80, second set of columns) to 0.62 (for age 85, first set of columns) to 1.23 (for age 90, third set of columns).

Thus, when answering the general questions of "does age matter" and "does grip strength matter" in predicting cognition, we might come to very different conclusions across versions of the same model due to differences in centering. But to resolve this apparent discrepancy, we must realize that because these effects moderate one another, the correct answer is "it depends" to both questions. In weaker persons (grip strength = 6), age *does* matter (significant simple main effect of age). In stronger persons (grip strength = 12), age *does not* matter (nonsignificant simple main effect of age). In younger persons (age = 80), grip strength *does not* matter (nonsignificant simple main effect of grip strength). In older persons (age = 90), grip strength *does* matter (significant simple main effect of grip strength).

2.D. Plotting Interactions Using Hypothetical People

Perhaps the easiest way to convey the pattern of any interaction effect is to create a figure that illustrates how the slopes of the predictors are moderated by each other. To do so, we will use the method of plotting interactions via **hypothetical people.**

In this approach, we first create fictional cases with targeted values for their predictor variables. We can then illustrate the pattern of the interaction by plotting the predicted outcomes for these hypothetical people as calculated from the fixed effects in the model and their values of the model predictors. Typically the predictors that are not part of the interaction to be plotted are held constant (i.e., sex and dementia group here). Furthermore, although it is common to use ±1 SD of the predictors as the plotted values, any meaningful value of the predictor can be used to generate predicted outcomes for the hypothetical people. Here we will create model-predicted cognition outcomes using the same predictor values as when we re-centered the predictors to examine simple slopes. The manual calculations needed to obtain each predicted outcome are shown in Equation (2.11):

$$
\begin{aligned}
&\text{Predicted Cognition}_i = \beta_0 + \beta_1(\text{Age}_i - 85) + \beta_2(\text{Grip}_i - 9) + \beta_6(\text{Age}_i - 85)(\text{Grip}_i - 9)\\
&\text{Grip Strength} = 12,\ \text{Age} = 80: \quad 29.41 - 0.33(-5) + 0.62(\ 3) + 0.12(-5)(\ 3) = 31.09\\
&\text{Grip Strength} = 12,\ \text{Age} = 85: \quad 29.41 - 0.33(\ 0) + 0.62(\ 3) + 0.12(\ 0)(\ 3) = 31.27\\
&\text{Grip Strength} = 12,\ \text{Age} = 90: \quad 29.41 - 0.33(\ 5) + 0.62(\ 3) + 0.12(\ 5)(\ 3) = 31.44\\
&\text{Grip Strength} = 9,\ \text{Age} = 80: \quad 29.41 - 0.33(-5) + 0.62(\ 0) + 0.12(-5)(\ 0) = 31.08\\
&\text{Grip Strength} = 9,\ \text{Age} = 85: \quad 29.41 - 0.33(\ 0) + 0.62(\ 0) + 0.12(\ 0)(\ 0) = 29.41\\
&\text{Grip Strength} = 9,\ \text{Age} = 90: \quad 29.41 - 0.33(\ 5) + 0.62(\ 0) + 0.12(\ 5)(\ 0) = 27.74\\
&\text{Grip Strength} = 6,\ \text{Age} = 80: \quad 29.41 - 0.33(-5) + 0.62(-3) + 0.12(-5)(-3) = 31.06\\
&\text{Grip Strength} = 6,\ \text{Age} = 85: \quad 29.41 - 0.33(\ 0) + 0.62(-3) + 0.12(\ 0)(-3) = 27.55\\
&\text{Grip Strength} = 6,\ \text{Age} = 90: \quad 29.41 - 0.33(\ 5) + 0.62(-3) + 0.12(\ 5)(-3) = 24.03
\end{aligned}
\tag{2.11}
$$

in which nine possible values are calculated (using the original fixed effect estimates) for each pairing of age (80, 85, or 90) and grip strength (6, 9, or 12). Note that the centered values of the predictors are included in Equation (2.11) rather than the original values. Also note that because the terms unrelated to age or grip strength (main effects of sex β_3 and dementia group β_4 and β_5) are not included, the predicted cognition outcomes are implicitly for a man ($\text{SexMW}_i = 0$) who will not be diagnosed with dementia ($\text{DemNF}_i = 0$ and $\text{DemNC}_i = 0$). Although using different values for these other predictors would adjust the intercept up or down, they would not adjust the simple main effects or interaction between age and grip strength, because sex and dementia group do not interact with age or grip strength in this model. Thus, we do not need to include the sex and dementia group variables to demonstrate the age by grip strength interaction, as now shown in Figure 2.1.

Figure 2.1 illustrates both interpretations of the two-way interaction between age and grip strength. First, using the top panel, consider the original version of the model in Equation (2.9), as shown in the first set of columns in Table 2.2, in which age was centered at 85 years and grip strength was centered at 9 pounds. The predicted intercept from that model of 29.41 is shown in the center point of the second line for age = 85 and grip strength = 9. The simple main effect of age is displayed via the slope of the second line (the difference in cognition across age for grip strength = 9 pounds). Similarly, the simple main effect of grip strength is displayed as the vertical distance between the lines at the middle points (the difference in cognition across grip strength for age 85). Still using the top panel of Figure 2.1, next consider

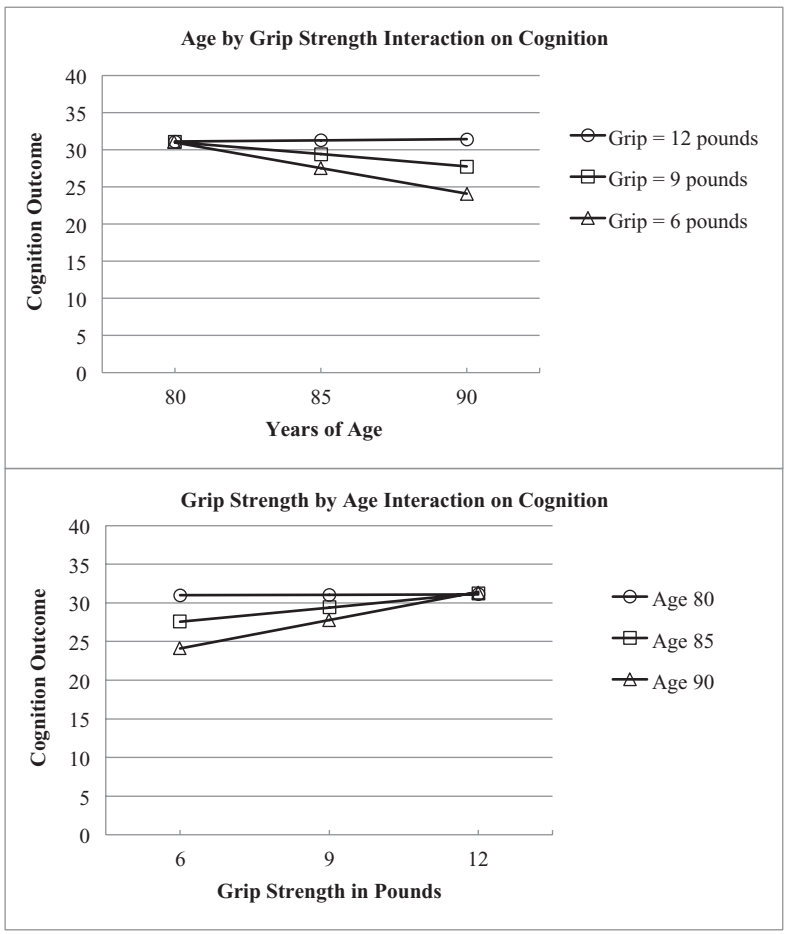

Figure 2.1 Decomposing an age by grip strength interaction via simple slopes for age (top) and simple slopes for grip strength (bottom).

the second reference point we created, as shown in the second set of columns in Table 2.2, in which age was centered at 80 years and grip strength was centered at 12 pounds. For that model, the intercept of 31.09 is shown as the left point on the top line, the simple main effect of age is shown by the slope of the top line, and the simple main effect of grip is shown by the vertical distance across the left points. Finally, still using the top panel of Figure 2.1, consider the third reference point we created, as shown in the third set of columns in Table 2.2, in which age was centered at 90 years and grip strength was centered at 6 pounds. For that model, the intercept of 24.03 is shown as the right point on the bottom line, the simple main effect of age is shown by the slope of the bottom line, and the simple main effect of grip is shown by the vertical distance across the right points.

To summarize, the top panel in Figure 2.1 shows how the main effect of age (as the slope of the lines) becomes less negative as grip strength is greater, as well as

how the main effect of grip strength (as the distance between the lines) becomes more positive in older persons. The bottom panel of Figure 2.1 then shows these same ideas but with grip strength on the *x*-axis instead, such that the slopes of the lines show the simple main effects of grip strength for each age, whereas the vertical distances between the lines then show the simple effects of age.

Although calculating the nine predicted outcomes to be plotted in Figure 2.1 was straightforward, such manual calculations (even via spreadsheets) can quickly become tedious and error prone. To automate the calculation of predicted outcomes to be plotted, we can take the concept of hypothetical people even further—by adding them to our dataset! More specifically, we would add to the data cases that carry all necessary combinations of the predictor values to be plotted. In this example, nine fake persons would be added to show every possible pairing of age (80, 85, or 90) and grip strength (6, 9, or 12). We could then take advantage of the fact that most software routines will provide model-predicted outcomes for each observation. Given that the hypothetical people do not have values for the outcome variable, their data will not be used in estimating the model, so no harm will come from including these hypothetical cases in the analysis. The model for the means then creates an expected outcome for all observations with values for the predictor variables. Thus one can easily obtain predicted outcomes to create plots for any person, hypothetical or real. Syntax for the creation and addition of hypothetical people and generation of predicted outcomes is given in the files for this example online.

2.E. Assessing Regions of Significance of Main Effects Within Interactions

So far we have examined how the interaction moderates each main effect (which then become simple main effects). We have also re-centered the predictors to get a sense of whether these simple main effects remain significant when evaluated at different values of the interacting predictor. For instance, we have learned that the effect of age is significantly negative when evaluated at the mean grip strength or 1 SD below the mean (grip strength = 9 and 6 pounds, respectively), but is nonsignificant when evaluated at 1 SD above the mean (grip strength = 12 pounds). Likewise, we have learned that the effect of grip strength is significantly positive when evaluated at age 90 or age 85, but is nonsignificant at age 80. A natural follow-up question, then, is across what range of the interacting predictor will a given simple main effect be significant?

Methods for decomposing an interaction effect via **regions of significance** (e.g., Johnson & Fay, 1950) have been extended more recently for use with interactions among continuous predictors (e.g., Bauer & Curran, 2005; Preacher, Curran, & Bauer, 2006). The idea behind *regions of significance* is this: Rather than picking arbitrary values of the interacting predictor at which to evaluate the significance of each simple main effect, we can instead determine the threshold values of the interacting predictor after which the simple main effect of the interacting predictor becomes

nonsignificant. For instance, rather than asking "is the effect of grip strength still significant at age 80?" we could instead ask "at what age does the grip strength effect become nonsignificant?" Alternatively, we could ask "at what level of grip strength does the age effect become nonsignificant?" In other words, assessing regions of significance allows us to obtain the points at which the simple slopes turn on, turn off, or even change direction (given that interaction effects imply nonparallel lines that will eventually cross).

As presented earlier in the chapter, the significance of a given fixed effect is determined by a Wald test: the *t*-statistic formed by the ratio of its estimated slope over its SE. If the *t*-statistic is smaller than –1.96 or greater than 1.96, the slope is deemed significant at the $\alpha = .05$ level. Thus, we can determine if a simple slope is significant for any value of the interacting predictor. To obtain the simple slopes that lie at the boundaries of significance, we need to turn this formula around—we must find the simple slopes that correspond to the desired *t*-statistics of ±1.96 (for slope / SE), and the values of the interacting predictor at which they occur.

To illustrate regions of significance using our current example, let us first determine the boundary ages at which the grip strength effect is no longer significant. To do so, we consider grip strength as the effect of interest and age as the moderator effect. Recall from Table 2.2 that the simple main effect of grip strength was nonsignificant at age 80 but was significantly positive at age 85. At what age does the effect of grip strength become significantly positive (i.e., when does the grip strength slope turn on)? To determine this, we need to consider which model effects are responsible for modifying the grip strength slope. Recall from Equation (2.10) that the simple slope for grip strength is a function of its simple main effect β_2 and its interaction with age β_6, such that grip strength slope = $\beta_2 + \beta_6(\text{Age}_i - 85)$. Thus, the grip strength slope SE is also a function of its simple main effect and interaction with age. The formulas to derive the simple slope estimate and its SE to compute the *t*-statistic for significance are shown in Equation (2.12):

$$
\begin{aligned}
&\pm t = \pm 1.96 = \frac{\text{Slope Estimate}}{\sqrt{\text{Variance of Slope Estimate}}}, \text{ where:} \\
&\text{Grip Strength Slope Estimate} = \beta_2 + \beta_6(\text{Age} - 85) \\
&\text{Variance of Slope Estimate} = \text{Var}(\beta_2) + 2\text{Cov}(\beta_2\beta_6)(\text{Age} - 85) \\
&\qquad\qquad + \text{Var}(\beta_6)(\text{Age} - 85)^2
\end{aligned}
\tag{2.12}
$$

in which Age – 85 is the value at which the simple slope of grip strength is to be evaluated. As shown in Equation (2.12), the total sampling variance of the grip strength slope estimate (i.e., its SE squared) is a complex function of the sampling variance of its own slope β_2 and the sampling variance of the interaction slope β_6. This latter equation is derived from general mathematical rules about how to find the expected variance of a random variable. In this case, the random variable is the grip strength slope estimate, computed as: $\beta_2 + \beta_6(\text{Age} - 85)$. Contributing to the total sampling variance of the slope estimate is not only the sampling variance of

each part, but also the covariance among the parts. Accordingly, in this case, the total sampling variance of the slope estimate includes the variance of the β_2 estimate, the variance of the β_6 estimate, and twice the covariance between them (given that a covariance matrix is symmetric and thus the covariance shows up twice). Because the interaction slope β_6 is a function of age, age is also included in the covariance with β_2. Finally, age is squared in the last term because it is treated as a constant, and constants can be removed from the expected variance if they are squared. Finally, through some tedious algebra we can solve for the slopes needed to obtain $t = \pm 1.96$ (see Bauer & Curran, 2005).

Although we can readily obtain the sampling variance of each estimated slope (i.e., as its SE squared), covariances among the estimates are generally not provided by default. However, one can request the **asymptotic covariance matrix**, which will contain the sampling variance of each estimated fixed effect as well as the covariances among the estimates of the fixed effects. The asymptotic covariance matrix is available only in some software procedures, which is another reason why we are using procedures for general linear mixed models to estimate these general linear models (as shown in the example syntax files online). Table 2.3 provides the asymptotic covariance matrix for the estimates of the fixed effects in the model in Equation (2.9) as well as the original fixed effect estimates when age is centered at 85 years and grip strength is centered at 9 pounds (bottom row). The values in bold are those needed to calculate regions of significance for the simple effect of grip strength as moderated by age, including the grip strength slope estimate β_2 and its sampling variance, the interaction slope estimate β_6 and its sampling variance, and the covariance between the β_2 and β_6 slope estimates.

Table 2.3 Covariance matrix for the Equation 2.9 parameter estimates. Bold values are used for assessing regions of significance.

Estimate Covariance	*Intercept* β_0	*Age* β_1	*Grip Strength* β_2	*Sex* β_3	*DemNF* β_4	*DemNC* β_5	*Age by Grip Strength* β_6
Intercept	0.4829						
Age	0.0005	0.0145					
Grip Strength	–0.0308	0.0033	**0.0221**				
Sex (0 = Men, 1 = Women)	–0.4507	0.0050	0.0537	0.7873			
DemNF (None vs. Future)	–0.1820	–0.0041	–0.0134	–0.0710	1.0274		
DemNC (None vs. Current)	–0.2263	–0.0012	–0.0003	0.0237	0.2129	2.2878	
Age by Grip Strength	0.0019	0.0010	**0.0002**	0.0027	–0.0027	0.0024	**0.0016**
Fixed Effect Estimate	29.4078	–0.3340	**0.6194**	–3.4556	–5.9225	–16.3004	**0.1230**

Calculators for computing regions of significance are available in the online resources; results may differ slightly depending on how many digits are carried forward into the calculations. When using all possible digits, the upper threshold for age is –2.29 (where $t = 1.96$) and the lower threshold for age is –14.82 (where $t = -1.96$). Given the centering of age as 0 = 85 years, these values translate (within rounding error) into –2.29 + 85 = 82.71 years and –14.82 + 85 = 70.18 years. Although the estimated values will differ if other centering points of age were used instead, the resulting thresholds will be the same, and can be interpreted as follows. Above age 82.71, there will be a significant *positive* effect of grip strength on cognition (i.e., as found when evaluating the effect of grip strength at age 85 or age 90 in Table 2.2). Between age 70.18 and 82.71, there will be a *nonsignificant* effect of grip strength on cognition (i.e., as found when evaluating the effect of grip strength at age 80 in Table 2.2). In addition, at age 70.18 or younger, there will be a significant *negative* effect of grip strength. Although this reversal of the effect of grip strength may seem strange, it is a natural consequence of any linear interaction effect—nonparallel lines eventually cross, so any positive effect must eventually become negative (and any negative effect must eventually become positive). In these data, the youngest person is 80 years old, and thus for most persons we would expect a positive effect of grip strength, which becomes even more positive in older persons via the age by grip strength interaction effect.

We can also assess the region of significance for the age slope as moderated by grip strength. In this case, we will need the age slope estimate β_1 and its sampling variance, the interaction slope estimate β_6 and its sampling variance, and the covariance between the β_1 and β_6 slope estimates. Using these values, the upper threshold for grip strength is 9.52 (where $t = 1.96$) and the lower threshold for grip strength is 0.67 (where $t = -1.96$). Given the centering of grip strength as 0 = 9 pounds, these values translate into 9.52 + 9 = 18.52 pounds and 0.67 + 9 = 9.67 pounds and can be interpreted as follows. Above 18.52 pounds of grip strength, there will be a significant *positive* effect of age on cognition. Between 9.67 and 18.52 pounds of grip strength, there will be a *nonsignificant* effect of age on cognition. Below 9.67 pounds of grip strength, there will be a significant *negative* effect of age on cognition (i.e., as found when evaluating the effect age at grip strength = 6 or 9 pounds in Table 2.2). Given that grip strength values in the current sample range from 0 to 19 with a mean of 9 pounds, for about half the sample we would expect to see a significant negative effect of age (which would become more negative as grip strength is weaker), while we would expect to see a minimal effect of age for the other upper half of the sample (and for almost no one would we expect to see a positive effect of age).

So far we have examined several tools to describe two-way interactions between continuous predictors. These include the process of translating marginal main effects into simple main effects when part of an interaction, re-centering predictors to obtain simple main effects at various points of interest, showing differences in simple slopes by plotting model-predicted outcomes for hypothetical people, and computing regions of significance to explore the point of the moderator at which the simple main effects turn on or turn off (and perhaps turn back on again in the opposite direction). Later in this chapter we will use some of these same tools to decompose three-way and higher-order interactions as well. For now, though, we

focus on expanding our repertoire to include interpretation of interactions among categorical predictors.

3. Interpreting Interactions Involving Categorical Predictors

To continue with our example, we can examine whether the differences in cognition that were found as a function of dementia diagnosis differ by sex. In this section the sex by dementia group interaction will be examined using each dementia group as the reference in turn.

The interaction of sex by dementia group is first specified using the no dementia group as the reference, as shown in Equation (2.13):

$$\begin{aligned}\text{Cognition}_i = {} & \beta_0 + \beta_1(\text{Age}_i - 85) + \beta_2(\text{Grip}_i - 9) + \beta_3(\text{SexMW}_i) \\ & + \beta_4(\text{DemNF}_i) + \beta_5(\text{DemNC}_i) + \beta_6(\text{Age}_i - 85)(\text{Grip}_i - 9) \\ & + \beta_7(\text{SexMW}_i)(\text{DemNF}_i) + \beta_8(\text{SexMW}_i)(\text{DemNC}_i) + e_i\end{aligned} \tag{2.13}$$

in which two new effects, β_7 and β_8, represent the interaction of sex with the three dementia groups (none, future, and current). Just as we needed two contrasts for the main effects of how the three dementia groups differ in cognition (β_4 and β_5), we need two contrasts to indicate how the three dementia groups differ in their effect of sex on cognition (or equivalently, to represent how dementia group differences manifest differently in women than in men). The fixed effects from this model are shown in the first set of columns in Table 2.4. Adding the β_7 and β_8 interaction terms further reduced the error variance in cognition to $\sigma_e^2 = 85.97$, for a total reduction from the original variation in cognition of approximately 29%, or $R^2 = (120.76 - 85.97) / 120.76 = .29$. Some of the effects from in the previous model with only main effects of sex and dementia group from Equation (2.9) now take on different interpretations due to the sex by dementia group interaction terms of β_7 and β_8. We will discuss each of these effects in turn. As in the previous section, small differences (i.e., ≤ 0.01) between the values calculated in the text and those reported in the tables may occur due to rounding error.

To begin, we note the terms in the model for the means that carry the same interpretation as in the previous model in Equation (2.9). Because no new predictors have been added to the model, the intercept β_0 is still the expected cognition outcome for an 85-year-old man with 9 pounds of grip strength who will not be diagnosed with dementia. Similarly, although the age slope β_1, the grip strength slope β_2, and the age by grip strength interaction β_6 are now the unique effects after also controlling for the sex by dementia group interaction, their interpretations do not change because they are not a part of the new interaction terms. Their obtained regions of significance are very similar to those found in the previous model. That is, the age slope will be significantly negative below a grip strength of 9.68 pounds, significantly positive above 18.65 pounds, and nonsignificant between 9.68 and 18.65 pounds. Likewise, the grip strength slope will be significantly negative below

Table 2.4 Fixed effects from models with sex by dementia interactions. Bold values are $p < .05$.

Model Parameters		*Equation 2.13 Reference = Men without Dementia*			*Equation 2.13 Reference = Women without Dementia*		
		Est	*SE*	*p <*	*Est*	*SE*	*p <*
β_0	Intercept	**29.07**	0.75	.001	**26.19**	0.64	.001
β_1	Age Slope (0 = 85 years)	**–0.33**	0.12	.005	**–0.33**	0.12	.005
β_2	Grip Strength Slope (0 = 9 lbs)	**0.62**	0.15	.001	**0.62**	0.15	.001
β_6	Age by Grip Interaction	**0.12**	0.04	.003	**0.12**	0.04	.003
β_3	Sex						
	Dementia: 0 = None	**–2.88**	1.01	.005	**2.88**	1.01	.005
	Dementia: 0 = Future						
	Dementia: 0 = Current						
	Dementia Group (None as Reference)						
β_4	DemNF: None vs. Future (0 = Men)	**–6.06**	1.64	.001			
β_4	DemNF: None vs. Future (0 = Women)				**–5.89**	1.28	.001
β_5	DemNC: None vs. Current (0 = Men)	**–11.97**	2.25	.001			
β_5	DemNC: None vs. Current (0 = Women)				**–19.85**	2.03	.001
	Dementia Group (Future as Reference)						
β_4	DemFN: Future vs. None (0 = Men)						
β_4	DemFN: Future vs. None (0 = Women)						
β_5	DemFC: Future vs. Current (0 = Men)						
β_5	DemFC: Future vs. Current (0 = Women)						
	Dementia Group (Current as Reference)						
β_4	DemCN: Current vs. None (0 = Men)						
β_4	DemCN: Current vs. None (0 = Women)						
β_5	DemCF: Current vs. Future (0 = Men)						
β_5	DemCF: Current vs. Future (0 = Women)						
	Sex by Dementia Group Interaction						
β_7	Sex by DemNF (None vs. Future)	0.16	2.07	.937	–0.16	2.07	.937
β_7	Sex by DemFN (Future vs. None)						
β_7	Sex by DemCN (Current vs. None)						
β_8	Sex by DemNC (None vs. Current)	**–7.88**	3.02	.010	**7.88**	3.02	.010
β_8	Sex by DemFC (Future vs. Current)						
β_8	Sex by DemCF (Current vs. Future)						

(*Continued*)

Table 2.4 (Continued)

Model Parameters		*Equation 2.15 Reference = Men with Future Dementia*			*Equation 2.15 Reference = Women with Future Dementia*		
		Est	*SE*	*p <*	*Est*	*SE*	*p <*
β_0	Intercept	**23.01**	1.49	.001	**20.30**	1.12	.001
β_1	Age Slope (0 = 85 years)	**–0.33**	0.12	.005	**–0.33**	0.12	.005
β_2	Grip Strength Slope (0 = 9 lbs)	**0.62**	0.15	.001	**0.62**	0.15	.001
β_6	Age by Grip Interaction	**0.12**	0.04	.003	**0.12**	0.04	.003
β_3	Sex						
	Dementia: 0 = None						
	Dementia: 0 = Future	–2.71	1.87	.149	2.71	1.87	.149
	Dementia: 0 = Current						
	Dementia Group (None as Reference)						
β_4	DemNF: None vs. Future (0 = Men)						
β_4	DemNF: None vs. Future (0 = Women)						
β_5	DemNC: None vs. Current (0 = Men)						
β_5	DemNC: None vs. Current (0 = Women)						
	Dementia Group (Future as Reference)						
β_4	DemFN: Future vs. None (0 = Men)	**6.06**	1.64	.001			
β_4	DemFN: Future vs. None (0 = Women)				**5.89**	1.28	.001
β_5	DemFC: Future vs. Current (0 = Men)	**–5.91**	2.59	.023			
β_5	DemFC: Future vs. Current (0 = Women)				**–13.95**	2.24	.001
	Dementia Group (Current as Reference)						
β_4	DemCN: Current vs. None (0 = Men)						
β_4	DemCN: Current vs. None (0 = Women)						
β_5	DemCF: Current vs. Future (0 = Men)						
β_5	DemCF: Current vs. Future (0 = Women)						
	Sex by Dementia Group Interaction						
β_7	Sex by DemNF (None vs. Future)						
β_7	Sex by DemFN (Future vs. None)	–0.16	2.07	.937	0.16	2.07	.937
β_7	Sex by DemCN (Current vs. None)						
β_8	Sex by DemNC (None vs. Current)						
β_8	Sex by DemFC (Future vs. Current)	**–8.04**	3.42	.019	**8.04**	3.42	.019
β_8	Sex by DemCF (Current vs. Future)						

Model Parameters		Equation 2.16 Reference = Men with Current Dementia			Equation 2.16 Reference = Women with Current Dementia		
		Est	*SE*	*p* <	*Est*	*SE*	*p* <
β_0	Intercept	**17.10**	**2.14**	**.001**	**6.35**	1.95	.001
β_1	Age Slope (0 = 85 years)	**–0.33**	0.12	.005	**–0.33**	0.12	.005
β_2	Grip Strength Slope (0 = 9 lbs)	**0.62**	0.15	.001	**0.62**	0.15	.001
β_6	Age by Grip Interaction	**0.12**	0.04	.003	**0.12**	0.04	.003
β_3	Sex						
	Dementia: 0 = None						
	Dementia: 0 = Future						
	Dementia: 0 = Current	**–10.75**	2.90	.001	**10.75**	2.90	.001
	Dementia Group (None as Reference)						
β_4	DemNF: None vs. Future (0 = Men)						
β_4	DemNF: None vs. Future (0 = Women)						
β_5	DemNC: None vs. Current (0 = Men)						
β_5	DemNC: None vs. Current (0 = Women)						
	Dementia Group (Future as Reference)						
β_4	DemFN: Future vs. None (0 = Men)						
β_4	DemFN: Future vs. None (0 = Women)						
β_5	DemFC: Future vs. Current (0 = Men)						
β_5	DemFC: Future vs. Current (0 = Women)						
	Dementia Group (Current as Reference)						
β_4	DemCN: Current vs. None (0 = Men)	**11.97**	2.25	.001			
β_4	DemCN: Current vs. None (0 = Women)				**19.85**	2.03	.001
β_5	DemCF: Current vs. Future (0 = Men)	**5.91**	2.59	.023			
β_5	DemCF: Current vs. Future (0 = Women)				**13.95**	2.24	.001
	Sex by Dementia Group Interaction						
β_7	Sex by DemNF (None vs. Future)						
β_7	Sex by DemFN (Future vs. None)						
β_7	Sex by DemCN (Current vs. None)	**7.88**	3.02	.010	**–7.88**	3.02	.010
β_8	Sex by DemNC (None vs. Current)						
β_8	Sex by DemFC (Future vs. Current)						
β_8	Sex by DemCF (Current vs. Future)	**8.04**	3.42	.019	**–8.04**	3.42	.019

70.06 years, significantly positive above 82.70 years, and nonsignificant between 70.06 and 82.70 years.

What have changed after adding the sex by dementia group interaction slopes β_7 and β_8 are the interpretations of the main effects for sex β_3 and dementia group β_4 and β_5. Previously, β_3 was the difference in cognition between men and women

(with men as the reference group), β_4 was the difference in cognition between the none and future dementia groups, and β_5 was the difference in cognition between the none and current dementia groups. Previously these effects were expected to hold equally across the sample (i.e., they were *unconditional* when they were not included in an interaction). But as seen by comparing the first set of columns in Table 2.2 to the first set of columns in Table 2.4, after adding their interactions, the main effect of sex β_3 changed from –3.46 to –2.88, the main effect of DemNF_i β_4 changed from –5.92 to –6.06, and the main effect of DemNC_i β_5 changed from –16.30 to –11.97.

These changes are again necessary and expected because, as we learned before, the main effects of an interaction become *conditional* on each other, such that they become the simple effects specifically when their interacting predictor is 0. To help us in interpreting these new simple main effects (as well as the interactions), we can use the model coefficients from the first set of columns in Table 2.4 to generate predicted values for all possible combinations of men and women in the three dementia groups, as shown in Equation (2.14):

$$
\begin{aligned}
\text{Predicted Cognition}_i &= \beta_0 + \beta_3(\text{SexMW}_i) + \beta_4(\text{DemNF}_i) + \beta_5(\text{DemNC}_i) \\
&\quad + \beta_7(\text{SexMW}_i)(\text{DemNF}_i) + \beta_8(\text{SexMW}_i)(\text{DemNC}_i) \\
\text{Men, None:} &\quad 29.07 - 2.88(0) - 6.06(0) - 11.97(0) + 0.16(0)(0) - 7.88(0)(0) = 29.07 \\
\text{Women, None:} &\quad 29.07 - 2.88(1) - 6.06(0) - 11.97(0) + 0.16(1)(0) - 7.88(1)(0) = 26.19 \\
\text{Men, Future:} &\quad 29.07 - 2.88(0) - 6.06(1) - 11.97(0) + 0.16(0)(1) - 7.88(0)(0) = 23.01 \\
\text{Women, Future:} &\quad 29.07 - 2.88(1) - 6.06(1) - 11.97(0) + 0.16(1)(1) - 7.88(1)(0) = 20.30 \\
\text{Men, Current:} &\quad 29.07 - 2.88(0) - 6.06(0) - 11.97(1) + 0.16(0)(0) - 7.88(0)(1) = 17.10 \\
\text{Women, Current:} &\quad 29.07 - 2.88(1) - 6.06(0) - 11.97(1) + 0.16(1)(0) - 7.88(1)(1) = 6.35
\end{aligned}
\tag{2.14}
$$

in which the intercept, main effects of sex and dementia group, and their two interactions are used to generate six predicted group means (each of which is assuming age = 85 years and grip strength = 9 pounds). Figure 2.2 illustrates these six group means as well. Now let us examine which of these group differences are given to us directly by the model in Equation (2.13), and which are not. Estimated group means are provided in parentheses to facilitate interpretation of the group mean differences represented by the coefficients, as described below.

Our reference point is men who will not be diagnosed with dementia ($\text{SexMW}_i = 0$, $\text{DemNF}_i = 0$, and $\text{DemNC}_i = 0$; the top left point in Figure 2.2). Accordingly, the simple main effect of sex β_3 is now the difference between men and women *specifically in the no dementia group* (the vertical distance between the left points in Figure 2.2), in which women without dementia (26.19) are predicted to have significantly lower cognition by 2.88 than men without dementia (29.07). Similarly, the simple main effect of DemNF_i β_4 is now conditional on men (the difference between the left and center points on the top line in Figure 2.2), such that men who will eventually be diagnosed with dementia (23.01) are expected to have significantly lower cognition by 6.06 than men who will not be diagnosed (29.07). Likewise, the simple main effect of DemNC_i β_5 is now conditional on men (the difference from the left to right

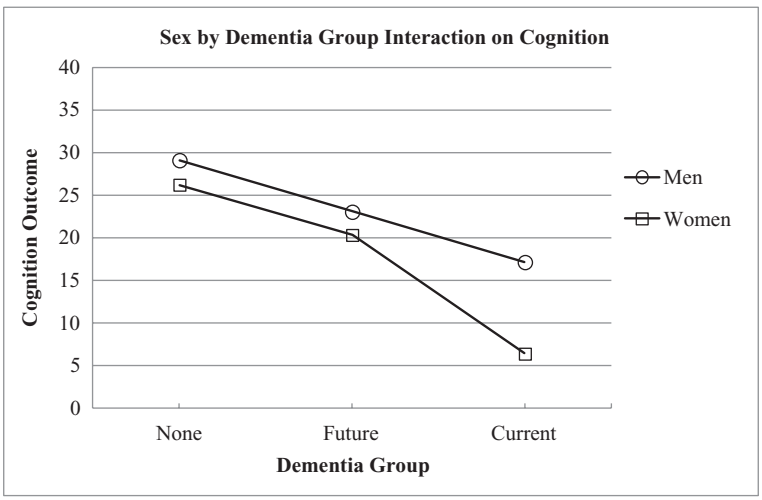

Figure 2.2 Decomposing a sex by dementia group interaction via simple slopes for dementia diagnosis group for each sex.

points on the top line in Figure 2.2), such that men who already have been diagnosed with dementia (17.10) are predicted to have significantly lower cognition by 11.97 than men who will not be diagnosed (29.07). These interpretations are based on the no dementia group as the reference (i.e., the group that has a value of 0 for both the $DemNF_i$ and $DemNC_i$ contrasts).

Turning to the two new interaction terms, we first consider the two possible (and equally correct) ways to interpret the nonsignificant interaction $\beta_7 = 0.16$ between $SexMW_i$ and $DemNF_i$. One way to interpret β_7 is as how the sex difference in cognition differs in the none versus future dementia groups. To do so, we start with the simple main effect of sex β_3: cognition in women and men without dementia differs by –2.88 (men = 29.07, women = 26.19 for no dementia). According to the interaction $\beta_7 = 0.16$, this sex difference of –2.88 becomes nonsignificantly less negative (smaller) by 0.16 in persons with future dementia ($DemNF_i = 1$), in which the sex difference is expected to be $\beta_3 + \beta_7 = -2.88 + 0.16 = -2.72$ (men = 23.01, women = 20.30 for future dementia). In other words, the nonsignificant interaction $\beta_7 = 0.16$ means that the sex difference (the vertical distance between the lines in Figure 2.2) is equivalent in the no dementia (–2.88; between left points) and future dementia groups (–2.72; between middle points).

The other way of interpreting the nonsignificant interaction $\beta_7 = 0.16$ is as how the difference in cognition between the none and future dementia groups differs by sex. To do so, we start with the simple main effect of $DemNF_i$ β_4: men with future dementia are expected to have lower cognition by 6.06 than men with no dementia (none = 29.07, future = 23.01 in men). The interaction $\beta_7 = 0.16$ tells us that the none versus future group difference of –6.06 in men becomes nonsignificantly less negative (smaller) by 0.16 in women, in which the none versus future group difference is expected to be $\beta_4 + \beta_7 = -6.06 + 0.16 = -5.91$ (none = 26.19, future = 20.30

in women). In other words, the nonsignificant interaction $\beta_7 = 0.16$ also means that the none versus future group difference (between the left and center points in Figure 2.2) is equivalent in men (–6.06; top line) and women (–5.91; bottom line).

Next, consider the two possible interpretations of the significant interaction $\beta_8 = -7.88$ between $SexMW_i$ and $DemNC_i$. Similar to the previous interaction β_7, one way to interpret the interaction β_8 is as how the sex difference in cognition differs in the none versus current dementia groups. To do so, we again start with the simple main effect of sex β_3: women and men without dementia differ by –2.88 (men = 29.07, women = 26.19 for no dementia). This sex difference of –2.88 becomes significantly more negative (larger) by 7.88 in persons currently with dementia ($DemNC_i = 1$), in which the sex difference is expected to be $\beta_3 + \beta_8 = -2.88 - 7.88 = -10.75$ (men = 17.10, women = 6.35 for current dementia). Thus, the significant interaction $\beta_8 = -7.88$ means that the sex difference favoring men (the vertical distance between the lines in Figure 2.2) is significantly larger in the current dementia group (–10.75; between right points) than in the no dementia group (–2.88; between left points).

The other way of interpreting the significant interaction β_8 is as how the difference in cognition between the none and current dementia groups differs by sex. To do so, we start with the simple main effect of $DemNC_i$ β_5: men with current dementia are expected to have lower cognition by 11.97 than men without dementia (none = 29.07, current = 17.10 in men). This none versus current group difference of –11.97 in men becomes significantly more negative (larger) by 7.88 in women, in which the none versus current group difference is expected to be $\beta_4 + \beta_8 = -11.97 - 7.88 = -19.85$ (none = 26.19, current = 6.35 in women). Thus, the significant interaction $\beta_8 = -7.88$ also means that the none versus current group difference (between the left and right points in Figure 2.2) is significantly larger in women (–19.85; bottom line) than in men (–11.97; top line).

Although the model in Equation (2.13) provides us with many possible contrasts among the six group means for sex by dementia group, it does not provide us with all of them. First, because men were used as the reference group, we now know whether two of the dementia group differences are significant for men (via the main effects of dementia group), and whether these dementia group differences are *different* for women (via their interactions with sex). But what is missing from our results is the explicit test of whether the implied dementia group differences for women are significant in and of themselves. To obtain the two simple main effects of dementia group specifically in women, we can change the reference group for sex to women by coding sex such that women = 0 and men = 1 instead. Results for this women-referenced model are shown in the second set of columns in Table 2.4, in which the intercept β_0 and the simple main effects of dementia group β_4 and β_5 are now specifically for women. Although the men-referenced model told us what these simple main effects for women should be (i.e., each is the simple main effect for men plus the interaction effect of how it differs for women), by using women as the reference group instead we can obtain the simple effect SEs and their corresponding *p*-values.

Accordingly, the none (26.19) versus future (20.30) dementia group difference $\beta_4 = -5.89$ in women (as could be calculated from the previous effect of $DemNF_i$ plus

how it differs in women of $\beta_4 + \beta_7 = -6.06 + 0.16 = -5.89$) is indeed significant, as is the none (26.19) versus current (6.35) group difference of –19.85 for women (as could be calculated from the previous effect of DemNC_i plus how it differs for women of $\beta_5 + \beta_8 = -11.97 - 7.88 = -19.85$). Because centering does not change the predictions of a model, the predicted outcomes from this women-referenced model will match those of the men-referenced model. Also, we see that any fixed effects that are unconditional with respect to sex (the age slope β_1, the grip strength slope β_2, and the age by grip interaction slope β_6) are the same as in the men-referenced model. Finally, we note that relative to the men-referenced model, the simple main effect of sex is exactly backwards, as are the interaction terms of sex by dementia group, as expected given the 0–1 switch in the coding for sex. Thus, the point of re-centering sex to make women the reference group is *not* to obtain a new simple main effect of sex or new interaction terms with sex, but to obtain new simple main effects for dementia group that test dementia group differences *specifically in women.*

In addition, our models thus far have included specific contrasts for the none versus current dementia groups and for the none versus future dementia groups, but we have not yet obtained explicit tests of the differences between the current and future dementia groups. To do so, we can change the coding of the dementia group predictors to make the future dementia group the reference instead of the no dementia group, as shown in Equation (2.15):

$$\begin{aligned}\text{Cognition}_i = {} & \beta_0 + \beta_1(\text{Age}_i - 85) + \beta_2(\text{Grip}_i - 9) + \beta_3(\text{SexMW}_i) \\ & + \beta_4(\text{DemFN}_i) + \beta_5(\text{DemFC}_i) + \beta_6(\text{Age}_i - 85)(\text{Grip}_i - 9) \\ & + \beta_7(\text{SexMW}_i)(\text{DemFN}_i) + \beta_8(\text{SexMW}_i)(\text{DemFC}_i) + e_i\end{aligned} \tag{2.15}$$

in which contrasts are now included for DemFN_i (none = 1, future = 0, current = 0) and DemFC_i (none = 0, future = 0, current = 1). Results from this model are shown in the third set of columns in Table 2.4. The results obtained for the simple main effect of future versus no dementia β_4 and its interaction with sex β_7 exactly mirror those found in the first set of columns in Table 2.4 from Equation (2.13), just in the opposite direction, as the reference group is now the future dementia group instead of the no dementia group. But new information is provided by the simple main effect of sex β_3, the simple main effect of future versus current dementia β_5, and their interaction β_8.

Specifically, the nonsignificant simple main effect of sex $\beta_3 = -2.71$ is now the difference between men and women *specifically in the future dementia group* (the vertical distance between the middle points in Figure 2.2). Although this effect could have been calculated from the model in Equation (2.13) (i.e., as the previous simple main effect of sex plus how it differs in the future dementia group of $\beta_3 + \beta_7 = -2.88 + 0.16 = -2.71$), these calculations would not provide an SE and *p*-value to assess its significance. This distinction turns out to be important, because although it is almost as large as was found in the no dementia group (–2.88, SE = 1.01), the estimated sex difference of –2.71 (SE = 1.87) in the future dementia group is nonsignificant. This is because there are fewer persons in future dementia group (20%)

than in the no dementia group (73%), and so the sex difference in the smaller future dementia group is estimated less precisely. Thus, although the interaction β_7 from the previous model in Equation (2.13) told us that the sex differences in cognition are equivalent in the none and future dementia groups, the sex difference is significant in the no dementia group only, likely due to differences in group sample size.

Additional new information is provided by the simple main effect of future versus current dementia ($DemFC_i$) $\beta_5 = -5.91$ (the difference between the middle and right points on the top line in Figure 2.2), indicating that men with current dementia (17.10) were predicted to have significantly lower cognition by 5.91 than men with future dementia (23.01). Finally, we also have the sex by future versus current dementia interaction $\beta_8 = -8.04$. This interaction indicates that the sex difference favoring men (the vertical difference between the lines in Figure 2.2) in the future dementia group (–2.71; between middle points) was significantly smaller than the sex difference in the current dementia group (–2.71 – 8.04 = –10.75; between right points). Or interpreted the other way, the difference between the future and current dementia groups (between the middle and right points in Figure 2.2) in men (–5.91; top line) was significantly smaller than in women (–5.91 – 8.04 = –13.95; bottom line).

As we did previously, we can re-estimate the model in Equation (2.15) using women as the reference group to obtain an SE and *p*-value for the future–current dementia group difference in women. As seen in the fourth set of columns in Table 2.4, the simple main effect of future versus current dementia $\beta_5 = -13.95$ in women (as could be calculated for the previous effect for men plus how it differs in women of $\beta_5 + \beta_8 = -5.91 - 8.04 = -13.95$) was also significant.

Let us now try to summarize our results, and see if we have missed any comparisons. First, consider the differences in cognition by dementia group within sex. We have learned that men and women with future or current dementia are predicted to have significantly lower cognition than men and women without dementia, as indicated by the simple main effects of $DemNF_i$/$DemFN_i$ and $DemNC_i$. Also, men and women with current dementia are predicted to have significantly lower cognition than men and women with future dementia, as indicated by the simple main effect of $DemFC_i$. The difference between the none and current dementia groups is the same for both sexes (as given by the nonsignificant sex by $DemNF_i$/$DemFN_i$ interaction), the difference between the none and future dementia groups is significantly greater for women (as given by the sex by $DemNC_i$ interaction), and the difference between the future and current dementia groups is also significantly greater for women (as given by the sex by $DemFC_i$ interaction).

Finally, let us consider the simple effects of sex within dementia group. In persons without dementia, men are expected to have significantly higher cognition than women (as given by the simple main effect of sex when none is the reference for dementia group). In persons with future dementia, men and women do not differ significantly (as given by the simple main effect of sex when future is the reference for dementia group). In persons with current dementia . . . this one is still missing! Working backwards from the previous model in Equation (2.15), we could calculate the sex difference for the current dementia group as the simple main effect of sex for future dementia plus the interaction for how it differs in for current

dementia, or $\beta_3 + \beta_8 = -2.71 - 8.04 = -10.75$. But is this sex effect significant, given that only 7% of the sample currently had dementia?

To obtain the missing simple main effect of sex, we can estimate one last set of models in which the current dementia group is the reference, as shown in Equation (2.16):

$$\begin{aligned}\text{Cognition}_i = {} & \beta_0 + \beta_1(\text{Age}_i - 85) + \beta_2(\text{Grip}_i - 9) + \beta_3(\text{SexMW}_i) \\ & + \beta_4(\text{DemCN}_i) + \beta_5(\text{DemCF}_i) + \beta_6(\text{Age}_i - 85)(\text{Grip}_i - 9) \\ & + \beta_7(\text{SexMW}_i)(\text{DemCN}_i) + \beta_8(\text{SexMW}_i)(\text{DemCF}_i) + e_i\end{aligned} \tag{2.16}$$

in which contrasts are now included for DemCN_i (none = 1, future = 0, current = 0) and DemCF_i (none = 0, future = 1, current = 0). Results from the model in Equation (2.16) are shown in the fifth set of columns in Table 2.4. The missing simple effect of sex for the current dementia group is given by $\beta_3 = -10.75$, which is significant. And although it provides no new information, the same model with women as the reference group instead of men is also reported for completeness in the sixth set of columns in Table 2.4. Thus, we can now conclude our summary of results with respect to sex differences in cognition (as given by the simple main effect of sex from each set of models): Men are predicted to have significantly higher cognition than women in the none and current dementia groups, but not in the future dementia group. Furthermore, the advantage for men is significantly greater in the current group than in the none group (by the sex by DemNC_i interaction) or than in the future group (by the sex by DemFC_i interaction), but the sex difference is equivalent in the none and future dementia groups (by the sex by DemNF_i interaction). Phew!

3.A. Requesting Simple Main Effects via Syntax From a Single Model

At this point you may notice all the redundancy in Table 2.4 and question whether all these re-centered versions of the same model are really necessary. The answer is both yes and no. Re-centering and re-estimating the same model multiple times may be necessary if you wish to obtain all possible simple main effects *directly from the fixed effects,* as we've done here. But as discussed earlier, this is not necessary if your software can provide estimates and standard errors for any fixed effect that is *implied* by the model (i.e., that is a linear combination of estimated fixed effects). In addition, by specifying predictors as "categorical" in the program syntax you can request all possible group means and comparisons among them directly (i.e., via LSMEANS in SAS, EMMEANS in SPSS, or MARGINS in STATA). This categorical predictor approach also often provides omnibus tests of whether the overall set of interactions is significant. In these data, the omnibus *F*-test of the sex by dementia group interaction (i.e., a multivariate Wald test) was $F(2, 541) = 3.49$, $p = .03$. From there, you can decompose the omnibus interaction into any specific group contrasts of interest—as we have already done the hard way by re-estimating the same model six times!

One important caveat (as illustrated in the appendix at the end of this chapter) is that you need to pay close attention to how the differences between groups are coded when interacting predictors are both specified as categorical. The programs will report group contrasts in which the main effects become *marginalized* over the interacting predictor (i.e., so that the main effect of sex is averaged across the three dementia groups rather than evaluated for the reference group, and so that the main effect of dementia group is averaged across men and women). Thus, these marginal main effects reported by the program may not agree with the simple (conditional) main effects obtained listed in the fixed effects—they shouldn't, because they mean different things.

Although they are both equally viable alternative representations of a main effect, a simple main effect may be more straightforward to interpret because it pertains to someone in a specific reference group that actually exists in the data. In contrast, because a marginal main effect (averaged across values of the interacting predictor) does not apply to a specific group, it may not be descriptive at any value of the interacting predictor. For instance, if a sex difference favoring men was found for one dementia group but a sex difference favoring women was found for another group, the *marginal* main effect of sex averaged across groups may be 0, because these two sex effects in different directions could cancel each other out. But the 0 marginal main effect of sex would not have any practical meaning in that case because it would not accurately describe the sex differences for any group. For this reason, in this text categorical predictors will be coded so that when they are included in an interaction, their main effects become conditional, simple effects (that refer to specific groups that actually exist in the data) rather than marginal effects (that refer to aggregate estimates created from combining across groups instead).

As discussed earlier in this chapter, when interpreting interactions among categorical predictors (e.g., as is typically done in ANOVA), summarizing the model via group means and their specific comparisons can be more convenient than describing the same model via simple main effects and specific interaction contrasts, as we've done instead here. But the purpose of this extended presentation was to demonstrate how we can estimate and interpret any kind of interaction within a general linear model, and not just those that are conveniently summarized via group means! Accordingly, in order to complete our interaction repertoire for use with more complex longitudinal models in later chapters, we now continue with an example of how to examine interactions between categorical and continuous predictors as well.

3.B. Interpreting Interactions Among Continuous and Categorical Predictors

Thus far we have examined whether cognition in older adults is related to age, grip strength, sex, and dementia diagnosis (none, future, or current). We have also examined whether the effect of age depends on grip strength (and vice-versa), as well as whether the effect of sex depends on dementia group (and vice-versa). To illustrate how to interpret interactions among a mix of categorical and continuous

predictors, we now examine whether the effect of sex depends on age or on grip strength (and vice-versa), as shown in Equation (2.17):

$$\begin{aligned}\text{Cognition}_i = {} & \beta_0 + \beta_1(\text{Age}_i - 85) + \beta_2(\text{Grip}_i - 9) + \beta_3(\text{SexMW}_i) \\ & + \beta_4(\text{DemNF}_i) + \beta_5(\text{DemNC}_i) + \beta_6(\text{Age}_i - 85)(\text{Grip}_i - 9) \\ & + \beta_7(\text{SexMW}_i)(\text{DemNF}_i) + \beta_8(\text{SexMW}_i)(\text{DemNC}_i) \\ & + \underline{\beta_9(\text{Age}_i - 85)(\text{SexMW}_i) + \beta_{10}(\text{Grip}_i - 9)(\text{SexMW}_i)} + e_i\end{aligned} \tag{2.17}$$

in which two new interactions of sex (using men as the reference) with age (centered at 85 years) and sex with grip strength (centered at 9 pounds) have now been included via β_9 and β_{10}, respectively. Results from the model in Equation (2.17) are shown in Table 2.5. Adding the β_9 and β_{10} interaction terms did not account for any additional variance in cognition ($R^2 = .30$, still); actually, the error variance from the model in Equation (2.17) was slightly larger than in the previous model ($\sigma_e^2 = 86.22$ vs. 85.97, previously). This strange result occurs because although the model sum of squares error term is indeed reduced slightly by the two new interactions, when divided by the residual degrees of freedom, the mean square error term is actually slightly higher than in the previous model. This kind of anomaly can happen for effects that are "really nonsignificant"—such as the two new interaction terms here. Nevertheless, we retain and interpret them for the sake of illustration and because they will be necessary to further augment the model to examine a three-way interaction in the next section.

Let us first consider how the simple effects change after adding interactions of age with sex (β_9) and grip strength with sex (β_{10}). Because of the age by sex interaction β_9, the significant simple main effect of age $\beta_1 = -0.39$ now applies specifically to men (as well as to grip strength of 9 pounds because of the previous age by grip strength interaction β_6). Because of the grip strength by sex interaction β_{10}, the main effect of grip strength $\beta_2 = 0.72$ is also now specific to men (as well as to an 85-year-old because of the previous age by grip strength interaction β_6). The significant simple main effect of sex $\beta_3 = -2.76$ is conditional on both new interactions—it now applies specifically to an 85-year-old with 9 pounds of grip strength (as well as to the no dementia group because of the previous sex by dementia group interactions β_7 and β_8).

There are two possible ways to interpret the nonsignificant age by sex interaction $\beta_9 = 0.08$, with each main effect serving as the moderator in turn. First, the significant age slope $\beta_1 = -0.39$ in men is nonsignificantly less negative by 0.08 in women (in which it would be $\beta_1 + \beta_9 = -0.39 + 0.08 = -0.31$, which was marginally significant, as shown in Table 2.5). Second, the significant advantage for men of $\beta_3 = -2.76$ found at age 85 narrows nonsignificantly by 0.08 per year of age (e.g., the advantage for men at age 86 would be $\beta_3 + \beta_9[\text{Age}_i - 85] = -2.76 + 0.08[1] = -2.68$). Although the simple effects of sex at other ages besides 85 are not shown in Table 2.5, in theory we could test the significance of any such alternative simple effects, or we could also examine regions of significance to determine at what ages the sex difference in cognition turns on or off. But these steps would be unnecessary here—the fact that the age by sex interaction β_9 is nonsignificant tells us that the

Table 2.5 Results from model including interactions of age by sex and grip strength by sex. Bold values are $p < .05$.

Model Parameters		*Equation 2.17*		
		Est	*SE*	*p <*
β_0	Intercept	**28.91**	0.80	.001
	Age Slope (0 = 85 years)			
β_1	Grip = 9, Men	**–0.39**	0.19	0.05
$\beta_1 + \beta_9$	Grip = 9, Women	–0.31	0.17	0.07
	Grip Strength Slope (0 = 9 lbs)			
β_2	Age = 85, Men	**0.72**	0.24	.002
$\beta_2 + \beta_{10}$	Age = 85, Women	**0.56**	0.19	.004
	Sex (0 = Men, 1 = Women)			
β_3	Grip = 9, Age = 85, No Dementia	**–2.76**	1.03	.008
$\beta_3 + \beta_7$	Grip = 9, Age = 85, Future Dementia	–2.53	1.90	.184
$\beta_3 + \beta_8$	Grip = 9, Age = 85, Current Dementia	**–10.64**	2.91	.001
	Dementia Group			
β_4	Men: None vs. Future	**–6.08**	1.64	.001
β_5	Men: None vs. Current	**–11.95**	2.25	.001
$\beta_5 - \beta_4$	Men: Future vs. Current	**–5.86**	2.59	.024
$\beta_4 + \beta_7$	Women: None vs. Future	**–5.86**	1.28	.001
$\beta_5 + \beta_8$	Women: None vs. Current	**–19.84**	2.03	.001
$\beta_5 + \beta_8 - \beta_4 - \beta_7$	Women: Future vs. Current	**–13.98**	2.24	.001
β_6	Age by Grip Interaction	**0.13**	0.05	.005
	Sex by Dementia Group Interaction			
β_7	Sex by None vs. Future	0.23	2.08	.913
β_8	Sex by None vs. Current	**–7.89**	3.03	.010
$\beta_8 - \beta_7$	Sex by Future vs. Current	**–8.12**	3.42	.018
β_9	Age by Sex Interaction	0.08	0.27	.774
β_{10}	Grip by Sex Interaction	–0.16	0.30	.590

effect of sex is the same across age. In addition, even if the age by sex interaction were significant, obtaining regions of significance for the age slope with respect to sex would not make any sense, because there are only two possible values of sex at which to evaluate of the age slope in predicting cognition anyway (as given in Table 2.5).

Let us now consider the two ways to interpret the other new (and also nonsignificant) interaction of grip strength by sex $\beta_{10} = –0.16$. First, the significant grip strength slope $\beta_2 = 0.72$ in men is nonsignificantly smaller (less positive) by 0.16 than in women (in which it would be $\beta_2 + \beta_{10} = 0.72 – 0.16 = 0.56$, which was still significant, as shown in Table 2.5). Second, the significant sex difference favoring

men $\beta_3 = -2.76$ at a grip strength of 9 pounds becomes nonsignificantly larger (more negative) by 0.16 for each additional pound of grip strength (e.g., the sex difference at a grip strength of 10 pounds would be $\beta_3 + \beta_{10}[\text{Grip}_i - 9] = -2.76 - 0.16[1] = -2.92$). It again would not make sense to determine the regions of grip strength for which the sex difference would remain significant because the nonsignificant grip strength by sex interaction tells us that the effect of grip strength on cognition is equivalent in men and women (and thus that the sex difference in cognition is the same across levels of grip strength as well).

Keeping track of the conditionality of the fixed effects in our current model is a little tricky, but we can do so as follows. The intercept β_0 is conditional on the 0 value for all model predictors (so age 85, grip strength of 9 pounds, men, no dementia group). Each main effect is then conditional on where its interacting predictors = 0, but not on the predictors with which it does not have an interaction. Thus, the age slope β_1 is conditional on grip strength of 9 pounds and men (but not on dementia group because age does not interact with dementia group). The grip strength slope β_2 is conditional on age 85 and men (but not on dementia group because grip strength does not interact with dementia group). The sex difference β_3 is conditional on age 85, grip strength of 9 pounds, and no dementia diagnosis. The dementia group differences for none versus future β_4 and none versus current β_5 are both conditional on men (but not on age or grip strength because they do not interact with age or grip strength). Finally, because the two-way interactions are the highest-order terms, they are unconditional. That is, the age by grip strength interaction is assumed constant over both sexes and dementia groups, the age by sex interaction is assumed constant over all grip strength and dementia groups, and the grip strength by sex interaction is assumed constant over all ages and dementia groups. To test these assumptions about the two-way interactions, we would need to estimate three-way interactions, as illustrated next.

3.C. Interpreting Three-Way and Higher-Order Interactions

Let us now consider how to interpret three-way (and higher-order) interactions. Although examples will also follow in later chapters, we will illustrate the general rules of interpreting higher-order interactions using our current example predicting cognition. For instance, let us examine a three-way interaction of age by grip strength by sex, as shown in Equation (2.18):

$$\begin{aligned}\text{Cognition}_i = {} & \beta_0 + \beta_1(\text{Age}_i - 85) + \beta_2(\text{Grip}_i - 9) + \beta_3(\text{SexMW}_i) \\ & + \beta_4(\text{DemNF}_i) + \beta_5(\text{DemNC}_i) + \beta_6(\text{Age}_i - 85)(\text{Grip}_i - 9) \\ & + \beta_7(\text{SexMW}_i)(\text{DemNF}_i) + \beta_8(\text{SexMW}_i)(\text{DemNC}_i) \\ & + \beta_9(\text{Age}_i - 85)(\text{SexMW}_i) + \beta_{10}(\text{Grip}_i - 9)(\text{SexMW}_i) \\ & + \underline{\beta_{11}(\text{Age}_i - 85)(\text{Grip}_i - 9)(\text{SexMW}_i)} + e_i\end{aligned} \tag{2.18}$$

the results for which are shown in Table 2.6. The three-way interaction $\beta_{11} = -0.16$ was not significant and did not account for any additional variance in cognition

(σ_e^2= 85.94; R^2 = .29). Nevertheless, we will retain and interpret it for the sake of illustration. But before we begin to decompose this model, note that in order for the three-way interaction to be interpreted correctly, all of its lower-order main effects and two-way interactions must be included in the model, regardless of their significance. In our case, this includes three main effects for age, grip strength, and

Table 2.6 Results from model including three-way interaction of age grip strength by sex. Bold values are p < .05.

Model Parameters		*Equation 2.18*		
		Est	*SE*	*p <*
β_0	Intercept	**28.96**	0.80	.001
	Age Slope (0 = 85 years)			
β_1	Grip = 9, Men	**–0.50**	0.21	.016
$\beta_1 + \beta_9$	Grip = 9, Women	**–0.40**	0.18	.026
	Grip Strength Slope (0 = 9 lbs)			
β_2	Age = 85, Men	**0.74**	0.24	.002
$\beta_2 + \beta_{10}$	Age = 85, Women	**0.54**	0.19	.005
	Sex (0 = Men, 1 = Women)			
β_3	Grip = 9, Age = 85, No Dementia	**–2.97**	1.04	.004
$\beta_3 + \beta_7$	Grip = 9, Age = 85, Future Dementia	–2.58	1.90	.175
$\beta_3 + \beta_8$	Grip = 9, Age = 85, Current Dementia	**–11.12**	2.92	.001
	Dementia Group			
β_4	Men: None vs. Future	**–6.17**	1.64	.001
β_5	Men: None vs. Current	**–11.78**	2.25	.001
$\beta_5 - \beta_4$	Men: Future vs. Current	**–5.62**	2.59	.031
$\beta_4 + \beta_7$	Women: None vs. Future	**–5.77**	1.28	.001
$\beta_5 + \beta_8$	Women: None vs. Current	**–19.93**	2.03	.001
$\beta_5 + \beta_8 - \beta_4 - \beta_7$	Women: Future vs. Current	**–14.16**	2.24	.001
	Age by Grip Interaction			
β_6	Men	**0.23**	0.08	.003
$\beta_6 + \beta_{11}$	Women	0.07	0.06	.214
	Sex by Dementia Group Interaction			
β_7	Sex by None vs. Future	**0.40**	2.08	.849
β_8	Sex by None vs. Current	**–8.15**	3.03	.007
$\beta_8 - \beta_7$	Sex by Future vs. Current	**–8.54**	3.43	.013
β_9	Age by Sex Interaction (for Grip = 9)	0.10	0.28	.729
β_{10}	Grip by Sex Interaction (for Age = 85)	–0.20	0.30	.512
β_{11}	Age by Grip by Sex Interaction	–0.16	0.10	.097

sex, as well as three two-way interactions of age by grip strength, age by sex, and grip strength by sex. Thus, even though the two latter two-way interactions were not significant, we must retain them to examine whether they depend on the third predictor (i.e., if the interaction of age by sex depends on grip strength, or if the interaction of sex by grip strength depends on age). Similarly, were we to estimate other three-way interactions (e.g., age by sex by dementia group), all of their lower-order main effects and two-way interactions would need to be included as well.

So what do we do with this three-way interaction? The rules are the same as when interpreting two-way interactions but are applied at a higher level of complexity. That is, just as two-way interactions modify their lower-order main effects (which then modify the intercept), three-way interactions modify their lower-order two-way interactions (which then modify their lower-order main effects, which modify the intercept). Furthermore, just as the main effects of an interaction become simple main effects specifically when the interacting predictor is 0, the two-way interactions within a three-way interaction become conditional on the third predictor = 0 as well. Thus, the simple two-way interaction of age by sex is specifically for grip strength = 0 (9 pounds), the simple two-way interaction of age by grip is specifically for sex = 0 (men), and the simple two-way interaction of sex by grip strength is specifically for age = 0 (85 years). Next we will examine all three possible interpretations of the three-way interaction. To do so, we will refer to Figure 2.3, which shows predicted cognition for a series of hypothetical people (men in the left panel, women in the right panel) who are age 80, 85, or 90 and who have grip strength of 6, 9, or 12 pounds (its mean and ± 1 SD).

First, we could describe how the two-way interaction of age by grip strength differs by sex. Previously we found that greater grip strength made the age slope less negative (and that older age made the grip strength slope more positive). Given the three-way interaction, the two-way interaction of age by grip strength $\beta_6 = 0.23$ is now specifically for men (the difference between the slope of the lines in the left

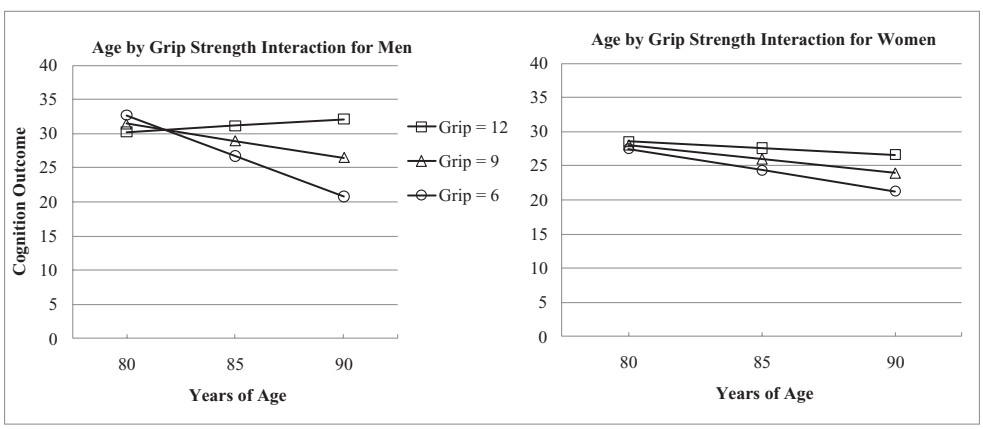

Figure 2.3 Decomposing an age by grip strength by sex interaction via simple slopes for age by grip strength for each sex.

panel of Figure 2.3). As reported in Table 2.6, the three-way interaction $\beta_{11} = -0.16$ indicates that the significant age by grip strength interaction $\beta_6 = 0.23$ in men is weaker (less positive) by 0.16 in women ($\beta_6 + \beta_{11} = 0.23 - 0.16 = 0.07$, which is not significant). This means age and grip strength depend on each other nonsignificantly more in men than women (i.e., the lines are more parallel in the right panel of Figure 2.3 for women).

Second, we could describe how the two-way interaction of age by sex differs for every additional pound of grip strength. Previously we found that the age slope was equivalent in men and women. Given the three-way interaction, the two-way interaction of age by sex $\beta_9 = 0.10$ is now specifically for someone with grip strength of 9 pounds (the non-difference in the slope of the middle line between men and women in Figure 2.3). The three-way interaction $\beta_{11} = -0.16$ indicates that the nonsignificant age by sex interaction $\beta_9 = 0.10$ at 9 pounds is nonsignificantly more negative by 0.16 for each additional pound of grip strength (so for someone with grip strength of 10 pounds, the age by sex interaction β_9 would be $\beta_9 + \beta_{11}[\text{Grip}_i - 9] = 0.10 - 0.16[1] = -0.06$). In other words, as grip strength is *lower,* the sex difference in the age slope becomes nonsignificantly *greater* (i.e., the difference between men and women in the slope of the bottom line in Figure 2.3 is greater than the difference between men and women in the slope of the other lines). If the three-way interaction were significant, we might also want to find the grip strength at which the sex difference in the age slope becomes significant (which would happen at some point below 9 pounds given the nonsignificant age by sex interaction $\beta_9 = 0.10$ at 9 pounds).

Third, we could describe how the two-way interaction of sex by grip strength differs for every additional year of age. Previously we found that the grip strength slope was equivalent in men and women. Given the three-way interaction, the two-way interaction of grip strength by sex $\beta_{10} = -0.20$ is now specifically for an 85-year-old (how the vertical difference between the middle points doesn't differ between men and women in Figure 2.3). The three-way interaction $\beta_{11} = -0.16$ indicates that the nonsignificant grip strength by sex interaction $\beta_{10} = -0.20$ for an 85-year-old is nonsignificantly more negative by 0.16 for each additional year of age (so for someone who is age 86, the grip strength by sex interaction β_{10} would be $\beta_{10} + \beta_{11}[\text{Age}_i - 85] = -0.20 - 0.16[1] = -0.36$). In other words, the nonsignificantly larger effect of grip strength in men than women is magnified nonsignificantly in older persons (the difference between men and women in the vertical difference between the lines in Figure 2.3 becomes larger in older ages). If the three-way interaction were significant, we might also want to find the age at which the sex difference in the grip strength slope becomes significant (somewhere past age 85, given the nonsignificant grip strength by sex interaction β_{10} at age 85).

Although the last model was quite complex, it could actually have been much worse, in that not all possible interactions were estimated! If we were feeling brave, we could try the two possible four-way interactions (age by grip by sex by none vs. future dementia; age by grip by sex by none vs. current dementia), which would each require two more two-way interactions (age by dementia, grip by dementia) and three more three-way interactions (age by grip by dementia; age by sex by

dementia; grip by sex by dementia). We could decompose the four-way interactions using the same basic strategies: we would discuss how a four-way interaction modifies each of its simple three-way interactions, which then modify their simple two-way interactions, which then modify their simple main effects (which then modify the intercept). Finally, because three-way and higher-order interactions can be more challenging to interpret, in addition to describing differences in simple main effects and interaction terms, plotting predicted outcomes created for hypothetical people with example values of the interacting predictors can be of great assistance. Such figures will be indispensable for the presentation of interaction results as well.

4. Chapter Summary

This chapter focused on between-person analysis via general linear models for predicting continuous outcomes (whose residuals should be conditionally normally distributed with constant variance and independence across persons). When each person has only one outcome and thus only one model residual, general linear models including one source of variation for differences between persons are useful for examining the effects of continuous (quantitative) or categorical (grouping) predictors, as well as interactions thereof. Traditionally, general linear models with continuous predictors are called *regression,* models with categorical predictors are called *analysis of variance,* and models with both kinds of predictors are called *analysis of covariance* or *regression.* Additional superficial differences between these models are found in how they are presented (i.e., via equations predicting individual outcomes or via tables of group mean outcomes), in the output provided by statistical programs (i.e., regression coefficients or cell mean differences), and in how main effects and interactions are specified (i.e., conditionally or marginally). Yet underneath all of these seemingly disparate models is a single general model that can be augmented to include any kind of predictor effect that is needed.

This chapter then tackled in great detail the potentially confusing world of interpreting interactions, beginning with interactions among continuous predictors, followed by interactions among categorical predictors, and then interactions among continuous and categorical predictors. Along the way we examined a new set of tools for decomposing any kind of interaction. First, you can always describe how the simple main effects are modified by their interaction terms (e.g., a positive main effect can become less positive or more positive; a negative main effect can become less negative or more negative). Second, you can change the 0 or conditional value of a predictor to evaluate the main effect of its interacting predictors at specific points of interest (e.g., to test the age slope specifically for men or women, to test the sex difference specifically among 80-year-olds or 90-year-olds), or accomplish the same goal by requesting fixed effects that are linear combinations of existing fixed effects in the program syntax. Third, you can calculate model-predicted outcomes for hypothetical people with prototypical values of the predictors in the interaction. These predicted outcomes can be created manually from the model equation, which can be useful for pedagogical purposes. Predicted outcomes can

also be calculated more quickly and with less error in three steps: (1) add the hypothetical people to your data set, (2) ask the program to create predicted values for each observation, and then (3) re-estimate the model. Predicted outcomes can then be plotted to more directly convey the pattern of the interaction. Finally, you can determine the regions along each moderator through which its interacting main effect is expected to be significant, or the values of the moderator at which the interacting main effect turns on or turns off. Calculating regions of significance can be especially useful when evaluating moderation by continuous predictors for which no particularly meaningful specific values exist. Furthermore, all four of these strategies can be used to decompose two-way, three-way, or even higher-order interactions.

In conclusion, let me say this: It is one thing to avoid retaining higher-order interaction terms because they are not statistically or practically significant; it is quite another to shy away from interactions because you don't know what they mean! Armed with a new (or revisited) set of tools for decomposing interactions, hopefully you will now approach any interaction effect with the confidence and clarity you've earned by making it all the way though this chapter!

5. Sample Results Section

The analyses in this chapter could be summarized into the beginning of a results section as follows (which would then need to be expanded to better capture the substantively meaningful story, theoretical framework, or set of research hypotheses to be tested). You'll note that instead of reporting by model parameter as we've done so far (i.e., all main effects, then all interactions), the text and table below are presented by predictor effect, which can be more intuitive to follow in models in which many of the main effects are conditional on higher-order interactions.

Between-person differences in cognition were examined in 550 older adults age 80 to 97 (M = 84.93 years, SD = 3.43). Cognition was measured by the Information Test, a measure of crystallized intelligence (M = 24.82, SD = 10.99, possible scores range from 0 to 44). The sample consisted of 41% men and 59% women. Other predictors included grip strength as measured in pounds per square inch (M = 9.11 pounds, SD = 2.99, range = 0 to 19 pounds) and dementia diagnosis group (none = 73%, future = 20%, or current = 7%). To facilitate interpretation of the intercept and main effects, each predictor was centered such that 0 was a meaningful value, including age (0 = 85 years), grip strength (0 = 9 pounds), and sex (0 = men, 1 = women). Finally, two contrasts were used represent differences among the three dementia diagnosis groups: DemNF (none = 0, future = 1, current = 0) and DemNC (none = 0, future = 0, current = 1). Main effects and interactions were added in sequential models. Significant effects were retained, as well as nonsignificant lower-order effects needed for significant interaction effects. Equation (2.13) provides the final model, the results of which are summarized in Table 2.7, Figure 2.1, and Figure 2.2. The significance of model parameters not directly given by Equation (2.13) was evaluated by requesting additional model-implied fixed effects.

Table 2.7 Example table of model results. Bold values are $p < .05$.

Model Effects		*Est*	*SE*	*p <*
Model for the Means				
β_0	Intercept	**29.07**	0.75	.001
β_1	Age Slope (0 = 85 years)	**–0.33**	0.12	.005
β_2	Grip Strength Slope (0 = 9 lbs)	**0.62**	0.15	.001
β_6	Age by Grip Interaction	**0.12**	0.04	.003
	Sex (0 = Men, 1 = Women) Differences			
β_3	No Dementia	**–2.88**	1.01	.005
$\beta_3 + \beta_7$	Future Dementia	–2.71	1.87	.149
$\beta_3 + \beta_8$	Current Dementia	**–10.75**	2.90	.001
	Dementia Group Differences			
	None vs. Future Dementia			
β_4	Men	**–6.06**	1.64	.001
$\beta_4 + \beta_7$	Women	**–5.89**	1.28	.001
β_7	Sex by None vs. Future	0.16	2.07	.937
	None vs. Current Dementia			
β_5	Men	**–11.97**	2.25	.001
$\beta_5 + \beta_8$	Women	**–19.85**	2.02	.001
β_8	Sex by None vs. Current	**–7.88**	3.02	.010
	Future vs. Current Dementia			
$\beta_5 - \beta_4$	Men	**–5.91**	2.59	.023
$\beta_5 + \beta_8 - \beta_4 - \beta_7$	Women	**–13.95**	2.24	.001
$\beta_8 - \beta_7$	Sex by Future vs. Current	**–8.04**	3.42	.019
Model for the Variance				
σ^2_e	Residual Variance	85.97		
	R^2 relative to Empty Model	.30		

The intercept $\beta_0 = 29.07$ is the expected cognition outcome for an 85-year-old man with 9 pounds of grip strength who will not be diagnosed with dementia later in the study. The main effect of age $\beta_1 = –0.33$ indicated that cognition is predicted to be significantly lower by 0.33 for every additional year of age (in persons with grip strength of 9 pounds). The main effect of grip strength $\beta_2 = 0.62$ indicated that cognition is predicted to be significantly greater by 0.62 for every additional pound of grip strength (in persons who are age 85). As shown in Figure 2.1, the age by grip strength interaction $\beta_6 = 0.12$ indicated the age slope predicting cognition became significantly less negative by 0.12 for each additional pound of grip strength (as shown by the differences in the slope of the lines). Equivalently, the grip

strength slope predicting cognition became significantly more positive by 0.12 for each additional year of age (as shown by the differences in the vertical distance between the lines).

To further decompose the age by grip strength interaction, the regions along each moderator through which the other main effect is expected to be significant were then calculated using the fixed effect estimates and their associated covariance matrix, as described in Bauer and Curran (2005). For the effect of age, the obtained threshold values of grip strength were 9.68 and 18.65 pounds. Given the range of grip strength of 0 to 19 pounds in the current sample (M = 9), the effect of age is expected to be negative for about half of the sample (below 9.68 pounds), the effect of age is expected to be nonsignificant for the other half (between 9.68 and 18.65 pounds), and the effect of age expected to be positive for almost no one (above 18.65 pounds). Similarly, for the effect of grip strength, the obtained threshold values of age were 70.06 and 82.70 years. Given the range of age of 80 to 97 years in the sample (M = 85), the effect of grip strength is expected to be negative for no one (below 70.06 years), the effect of grip strength is expected to be nonsignificant for a small part of the sample (between 70.06 and 82.70 years), and the effect of grip strength is expected to be positive for the majority of the sample (above 82.70 years).

The main and interactive effects of sex by dementia diagnosis group are presented next, as illustrated in Figure 2.2, in which the sex differences are shown by the vertical distances between the lines, and the diagnosis group differences are shown by the differences within the lines. First, with respect to sex differences, there was a significant main effect of sex $\beta_3 = -2.88$ such that in the no dementia group, cognition was significantly lower by 2.88 in women than in men. The sex difference in cognition was equivalent in no dementia and future dementia groups, as shown by the nonsignificant sex by no dementia versus future dementia interaction $\beta_7 = 0.16$. However, the resulting sex difference in cognition favoring men in the future dementia group of $\beta_3 + \beta_7 = -2.88 + 0.16 = -2.71$ was not significant, likely a result of the small number of persons with future dementia (only 20% of the sample). In addition, the sex difference in cognition was significantly larger in the current dementia group than in the no dementia group, as shown by the significant sex by no dementia versus current dementia interaction $\beta_8 = -7.88$, and the resulting sex difference in the current dementia group of $\beta_3 + \beta_8 = 2.88 - 7.88 = -10.75$ was also significant. The sex difference in cognition was also significantly larger in the current dementia group than in the future dementia group, as found by $\beta_8 - \beta_7 = -7.88 - 0.16 = -8.04$.

Next, with respect to differences among the dementia groups, cognition was significantly lower in the future dementia than no dementia group both in men, $\beta_4 = -6.06$, and in women, $\beta_4 + \beta_7 = -6.06 + 0.16 = -5.89$. This group difference was equivalent across sexes, as indicated by the nonsignificant sex by no dementia versus future dementia interaction $\beta_4 = 0.16$. Cognition was also significantly lower in the current dementia than no dementia group both in men, $\beta_5 = -11.97$, and in women, $\beta_5 + \beta_8 = -11.97 - 7.88 = -19.85$. This group difference was significantly larger in women, as indicated by the sex by no dementia versus current dementia interaction $\beta_8 = -7.88$. Finally, cognition was also significantly lower in the current dementia group than future diagnosis group both in men, $\beta_5 - \beta_4 = -11.97 + 6.06 = -5.91$, and

in women, $\beta_5 + \beta_8 - \beta_4 - \beta_7 = -11.97 - 7.88 + 6.06 + 0.16 = -13.95$. This group difference was significantly larger in women, as indicated by the additional interaction contrast of $\beta_8 - \beta_7 = -7.88 - 0.16 = -8.04$.

Review Questions

1. What makes a model *between-person?* Why might you predict that between-person models (i.e., general linear models) will not be suitable for longitudinal data?
2. Describe the process of centering continuous predictors. Which model effects should change as a result of choosing a different centering point, and which should not?
3. Describe the process of centering categorical predictors. What are the interpretational advantages and disadvantages of creating marginal vs. conditional (simple) contrasts?
4. Create your own examples to describe how a two-way interaction effect would be interpreted between all possible combinations of continuous with categorical predictors.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and probing interactions*. Thousand Oaks, CA: Sage.

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research, 40,* 373–400.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.

Johnson, P. O., & Fay, L. C. (1950). The Johnson-Neyman technique, its theory and application. *Psychometrika, 15,* 349–367.

Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data*. Mahwah, NJ: Erlbaum.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31,* 437–448.

Appendix 2.A: Marginal Versus Simple Main Effects in General Linear Modeling Output

As introduced earlier in the chapter, there are two choices as to how to specify fixed effects of categorical predictors in a general linear model (as well as in the general linear mixed model or multilevel models we'll see later). One option is for the user to code the differences between groups manually as separate variables (e.g., by creating the $DemNF_i$ and $DemNC_i$ dummy codes for the contrasts among the three groups). The other option is to let the software program create group contrasts by denoting the predictor as "categorical" (e.g., via the CLASS statement in SAS,

the BY statement in SPSS, or by using the i. indicator in STATA). Although neither approach is more correct than the other, you should be aware of how this choice will change the output that is then provided and what the resulting significance tests will then mean. Note that the following discussion does not apply to slopes for continuous variables, however.

Table 2.8 summarizes the output from Equation (2.13) when manually creating contrasts for dementia group in which the no dementia group is as the reference (i.e., $DemNF_i$ and $DemNC_i$, in which none = 0). Sex is represented by a single contrast between men (0) and women (1). The results under the heading "Parameter Estimates," provided by requesting the "solution for fixed effects" or "parameter estimates solution," for each fixed effect include an estimate, standard error, a *t*-statistic, and a *p*-value for the *t*-statistic (using denominator degrees of freedom = 541). For ease of comparison with the ANOVA solution, I have also computed the corresponding *F*-statistic (given as t^2 because each effect is tested using 1 degree of freedom in the numerator). The results under the heading "ANOVA" are provided by the Type III tests of the fixed effects in the ANOVA solution, which provides *F*-statistics instead of *t*-statistics. What Table 2.8 shows is the exact same information is provided by both output tables—the *F*-statistics match within rounding error, and the *p*-values match exactly. Furthermore, the interpretation of the estimates is based on the manual coding—in this case, men without dementia were the reference group for each of their simple main effects.

Table 2.9 provides the results from Equation (2.13) when sex and dementia group are specified as *categorical* predictors instead, such that the program then creates the contrasts instead of the user. First, notice that for each of the main effects of the categorical predictors (sex and dementia group), one row of output lists an estimate of 0.00 with dots where the standard errors and significance tests should be, which denotes the level of the grouping predictor that is serving as the reference category. Here, the highest coded group (or last alphabetically) is the reference, as

Table 2.8 Fixed effects significance tests given by the solution for fixed effects and analysis of variance (ANOVA) results when denoting sex and dementia group as continuous variables using manually coded contrasts. Bold values are $p < .05$.

Fixed Effects	*Parameter Estimates*					*ANOVA*	
	Est	*SE*	*t-value*	*F-value*	*p <*	*F-value*	*p <*
Intercept	**29.07**	0.75	38.84	1508.55	.001		
Age (0 = 85 years)	**–0.33**	0.12	–2.79	7.78	.005	7.80	.005
Grip Strength (0 = 9 lbs)	**0.62**	0.15	4.17	17.39	.001	17.41	.001
Age by Grip Strength	**0.12**	0.04	3.03	9.18	.003	9.16	.003
Sex (0 = Men, 1 = Women)	**–2.88**	1.01	–2.84	8.07	.005	8.09	.005
Dementia: None vs. Future	**–6.06**	1.64	–3.70	13.69	.000	13.72	.000
Dementia: None vs. Current	**–11.97**	2.25	–5.33	28.41	.001	28.43	.001
Sex by Dementia: None vs. Current	0.16	2.07	0.08	0.01	.937	0.01	.937
Sex by Dementia: None vs. Future	**–7.88**	3.02	–2.60	6.76	.010	6.78	.010

Table 2.9 Fixed effects significance tests given by the solution for fixed effects and analysis of variance (ANOVA) results when denoting sex and dementia group as categorical variables instead of using manually coded contrasts. Bold values are $p < .05$.

Fixed Effects	*Categorical Variables*		*Parameter Estimates*					*ANOVA*	
	Sex	*Dementia*	*Est*	*SE*	*t-value*	*F-value*	*p <*	*F-value*	*p <*
Intercept			**6.35**	1.95	3.26	10.63	.001		
Age (0 = 85 years)			**–0.33**	0.12	–2.79	7.78	.005	7.80	.005
Grip Strength (0 = 9 lbs)			**0.62**	0.15	4.17	17.39	.001	17.41	.001
Age by Grip Strength			**0.12**	0.04	3.03	9.18	.003	9.16	.003
Sex (0 = Men, 1 = Women)	Men		**10.75**	2.90	3.71	13.76	.000		
	Women		0.00	.	.		.	19.45	.001
Dementia Group (None, Future, or Current)		None	**19.85**	2.03	9.78	95.65	.001		
		Future	**13.95**	2.24	6.23	38.81	.001		
		Current	0.00	.	.		.	64.62	.001
Sex by Dementia Group	Men	None	**–7.88**	3.02	–2.6	6.76	.010		
	Men	Future	**–8.04**	3.42	–2.35	5.52	.019		
	Men	Current	0.00	.	.		.		
	Women	None	0.00	.	.		.		
	Women	Future	0.00	.	.		.		
	Women	Current	0.00	.	.		.	3.49	.031

in SAS or SPSS by default, although the lowest is the default in STATA (although this can be changed). Thus, for the effect of sex (a variable in which men = 0 and women = 1 in the dataset), the program essentially codes it backwards, such that women = 0 and men = 1. Similarly, for dementia group (a variable in which 1 = none, 2 = future, and 3 = current in the dataset), the highest-coded group, current dementia, becomes the reference instead.

This means that, in contrast to how the variables are actually coded in the dataset, the reference group in the parameter estimates solution for the intercept is a women with current dementia, the simple main effect of sex is the difference between women and men specifically in the current dementia group, and the simple main effects of dementia group are the differences specifically in women between current dementia and no dementia or between current dementia and future dementia. Likewise, the sex by dementia group interaction has four rows of 0.00 estimates and dots, but provides interaction effects for how the sex difference reported for the current dementia group differs in the no dementia or future dementia groups. Thus, the program has overridden the reference groups originally created by the user in

estimating the categorical predictors. Further complicating our interpretation is the fact that the *F*-statistics and significance tests thereof for the effects of the categorical predictors do not match across the parameter estimates solution and the ANOVA results. This is due to two factors.

First, for the main effects, whereas the *F*-statistics calculated in the parameter estimates solution are assessing *simple main effects,* the *F*-statistics provided in the ANOVA solution are assessing *marginal main effects* instead. For instance, the *F*-statistic for the main effect of sex in the parameter estimates (13.76) is testing the sex difference in the current dementia group specifically, whereas the *F*-statistic in the ANOVA solution (19.45) is testing the sex difference on average across the dementia groups. Thus, these *F*-statistics for the main effect of sex have different values because they mean different things (simple versus marginal main effects).

Second, for categorical predictors with more than two groups, the *F*-statistics also do not have the same degrees of freedom. As a result, the *F*-statistics for the main effect of dementia group do not correspond at all. Whereas the *F*-statistics in the parameter estimates solution are testing simple main effects for each group contrast with one degree of freedom (current vs. none, current vs. future, both specifically in women), the *F*-statistic for the main effect of dementia group in the ANOVA solution is testing the overall (i.e., "omnibus") difference across the groups using two degrees of freedom, *and* is testing those omnibus dementia group differences averaged across sex. Thus, the *F*-statistics for the main effect of group (two values in the parameter estimates solution; one value in the ANOVA solution) are providing completely different information. The same is true for the sex by dementia group interaction: the two single degree of freedom *F*-statistics in the parameter estimates solution are testing how the sex effect found in the current dementia group differs in each of the other groups, whereas the single *F*-statistic with two degrees of freedom in the ANOVA solution instead is testing the overall or omnibus interaction between the two levels of sex and the three levels of dementia group.

But specifying categorical predictors with the program rather than creating manual contrasts can still be advantageous despite the interpretational challenges the resulting output can present. Letting the program do the coding can be much easier than determining the exact coding scheme needed to represent complex interaction effects. It also makes it easier for the user to follow-up on an omnibus effect to request any additional group comparisons of interest, as well as to determine if a set of contrasts representing an overall omnibus interaction (e.g., as in sex by dementia group) is significant (rather than each contrast individually).

Finally, given an interaction effect, which should be reported, simple or marginal main effects? I believe that you as the analyst should decide which effects correspond more closely to your research questions, and report the effects that answer those questions accordingly. I personally find little use in marginal main effects when an interaction is present, because marginal main effects do not apply to any specific person or group, and thus may not describe anyone in the sample. However, there may be situations in which the main effect averaged over other interacting predictors can be useful. The main thing to remember in either case, though, is to be absolutely clear about what a given effect represents, and not just whether it is significant!