Raudenbush, S.W., & Bryk, A.S. (2002). Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd ed. Thousand Oaks: SAGE.

2 The Logic of Hierarchical Linear Models

- Preliminaries
- A General Model and Simpler Submodels
- Generalizations of the Basic Hierarchical Linear Model
- Choosing the Location of *X* and *W* (*Centering*)
- Summary of Terms and Notation Introduced in This Chapter

This chapter introduces the logic of hierarchical linear models. We begin with a simple example that builds upon the reader's understanding of familiar ideas from regression and analysis of variance (ANOVA). We show how these common statistical models can be viewed as special cases of the hierarchical linear model. The chapter concludes with a summary of some definitions and notation that are used throughout the book.

Preliminaries

A Study of the SES-Achievement Relationship in One School

We begin by considering the relationship between a single student-level predictor variable (say, socioeconomic status [SES]) and one student-level outcome variable (mathematics achievement) within a single, hypothetical school. Figure 2.1 provides a scatterplot of this relationship. The scatter of points is well represented by a straight line with intercept β_0 and slope β_1. Thus, the regression equation for the data is

$$Y_i = \beta_0 + \beta_1 X_i + r_i. \qquad [2.1]$$

The intercept, β_0, is defined as the expected math achievement of a student whose SES is zero. The slope, β_1, is the expected change in math achievement associated with a unit increase in SES. The error term, r_i, rep-

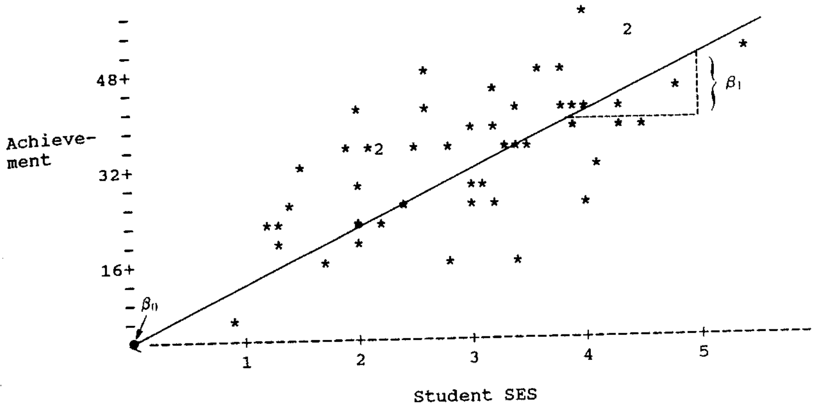

Figure 2.1. Scatterplot Showing the Relationship Between Achievement and SES in One Hypothetical School

resents a unique effect associated with person i. Typically, we assume that r_i is normally distributed with a mean of zero and variance σ^2, that is, $r_i \sim \mathrm{N}(0, \sigma^2)$.

It is often helpful to scale the independent variable, X, so that the intercept will be meaningful. For example, suppose we "center" SES by subtracting the mean SES from each score: $X_i - \overline{X}.$ where $\overline{X}.$ is the mean SES in the school. If we now plot Y_i as a function of $X_i - \overline{X}.$ (see Figure 2.2) with the

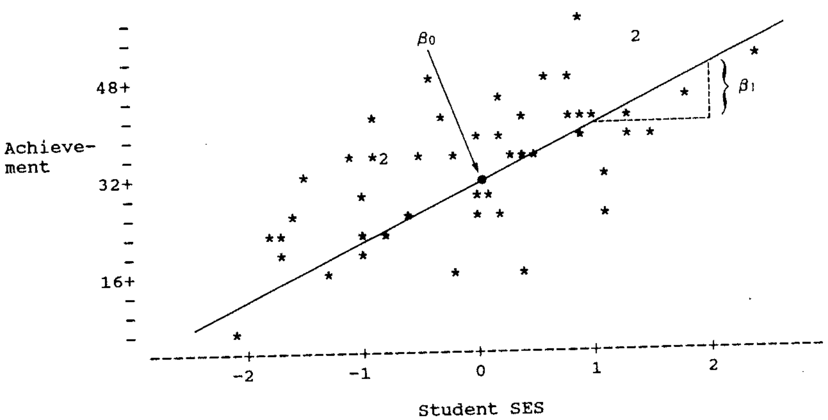

Figure 2.2. Scatterplot Showing the Relationship Between Achievement and SES (Centered) in One Hypothetical School

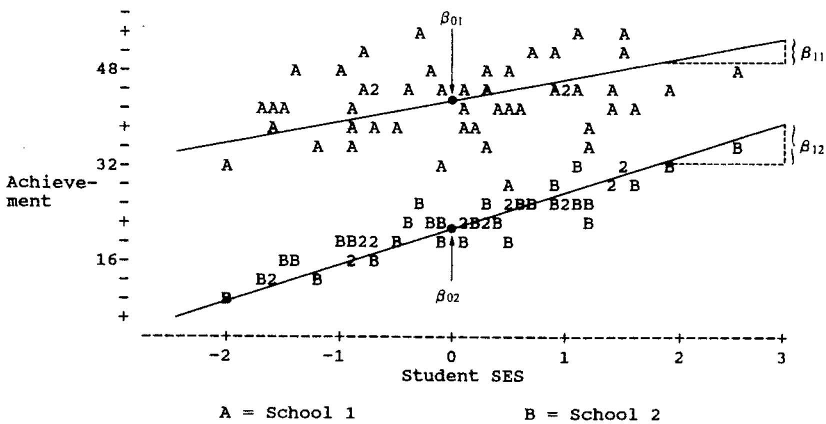

Figure 2.3. Scatterplot Showing the Relationship Between Achievement and SES Within Two Hypothetical Schools

regression line superimposed, we see that the intercept, β_0, is now the mean math achievement while the slope remains unchanged.

A Study of the SES-Achievement Relationship in Two Schools

Let us now consider separate regressions for two hypothetical schools. These are displayed in Figure 2.3. The two lines indicate that School 1 and School 2 differ in two ways. First, School 1 has a higher mean than does School 2. This difference is reflected in the two intercepts, that is, $\beta_{01} > \beta_{02}$. Second, SES is less predictive of math achievement in School 1 than in School 2, as indicated by comparing the two slopes, that is, $\beta_{11} < \beta_{12}$.

If students had been randomly assigned to the two schools, we could say that School 1 is both more "effective" and more "equitable" than School 2. The greater effectiveness is indicated by the higher mean level of achievement in School 1 (i.e., $\beta_{01} > \beta_{02}$). The greater equity is indicated by the weaker slope (i.e., $\beta_{11} < \beta_{12}$). Of course, students are not typically assigned at random to schools, so much interpretations of school effects are unwarranted without taking into account differences in student composition. Nevertheless, the assumption of random assignment clarifies the goals of the analysis and simplifies our presentation.

A Study of the SES-Achievement Relationship in *J* Schools

We now consider the study of the SES-math achievement relationship within an entire *population* of schools. Suppose that we now have a random sample of J schools from a population, where J is a large number. It is no longer practical to summarize the data with a scatterplot for each school. Nevertheless, we can describe this relationship within any school j by the equation

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{.j}) + r_{ij}, \qquad [2.2]$$

where for simplicity we assume that r_{ij} is normally distributed with homogeneous variance across schools, that is, $r_{ij} \sim \mathrm{N}(0, \sigma^2)$. Notice that the intercept and slope are now subscripted by j, which allows each school to have a unique intercept and slope. For each school, effectiveness and equity are described by the pair of values (β_{0j}, β_{1j}). It is often sensible and convenient to assume that the intercept and slope have a bivariate normal distribution across the population of schools. Let

$$\mathrm{E}(\beta_{0j}) = \gamma_0, \qquad \mathrm{Var}(\beta_{0j}) = \tau_{00},$$

$$\mathrm{E}(\beta_{1j}) = \gamma_1, \qquad \mathrm{Var}(\beta_{1j}) = \tau_{11},$$

$$\mathrm{Cov}(\beta_{0j}, \beta_{1j}) = \tau_{01},$$

where

γ_0 is the average school mean for the population of schools;
τ_{00} is the population variance among the school means;
γ_1 is the average SES-achievement slope for the population;
τ_{11} is the population variance among the slopes; and
τ_{01} is the population covariance between slopes and intercepts.

A positive value of τ_{01} implies that schools with high means tend also to have positive slopes. Knowledge of these variances and of the covariance leads directly to a formula for calculating the population correlation between the means and slopes:

$$\rho(\beta_{0j}, \beta_{1j}) = \tau_{01}/(\tau_{00}\,\tau_{11})^{1/2}. \qquad [2.3]$$

In reality, we rarely know the true values of the population parameters we have introduced $(\gamma_0, \gamma_1, \tau_{11}, \tau_{00}, \tau_{01})$ nor of the true individual school means and slopes (β_{0j} and β_{1j}). Rather, all of these must be estimated from the data. Our focus in this chapter is simply to clarify the meaning of the parameters. The actual procedures used to estimate them are introduced in Chapter 3 and are discussed more extensively in Chapter 14.

Suppose we did know the true values of the means and slopes for each school. Figure 2.4 provides a scatterplot of the relationship between β_{0j} and

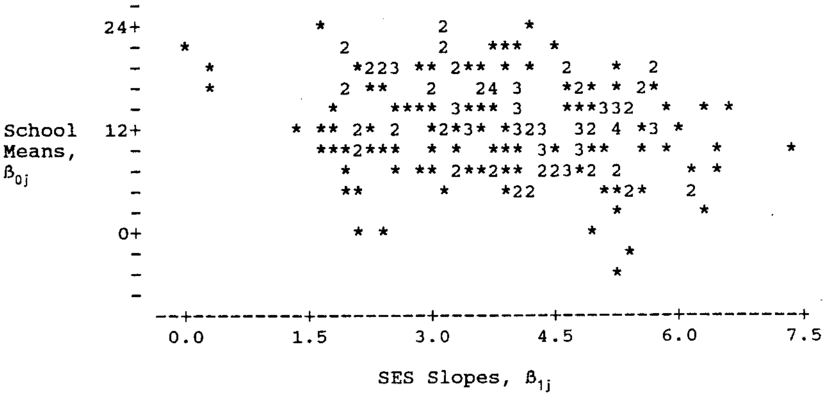

Figure 2.4. Plot of School Means (vertical axis) and SES Slopes (horizontal axis) for 200 Hypothetical Schools

β_{1j} for a hypothetical sample of schools. This plot tells us about how schools vary in terms of their means and slopes. Notice, for example, that there is more dispersion among the means (vertical axis) than the slopes (horizontal axis). Symbolically, this implies that $\tau_{00} > \tau_{11}$. Notice also that the two effects tend to be negatively correlated: Schools with high average achievement, β_{0j}, tend to have weak SES-achievement relationships, β_{1j}. Symbolically, $\tau_{01} < 0$. Schools that are effective and egalitarian—that is, with high average achievement (large values of β_{0j}) and weak SES effects (small values of β_{1j})—are found in the upper left quadrant of the scatterplot.

Having examined graphically how schools vary in terms of their intercepts and slopes, we may wish to develop a model to predict β_{0j} and β_{1j}. Specifically, we could use school characteristics (e.g., levels of funding, organizational features, policies) to predict effectiveness and equity. For instance, consider a simple indicator variable, W_j, which takes on a value of one for Catholic schools and a value of zero for public schools. Coleman, Hoffer, and Kilgore (1982) argued that W_j is positively related to effectiveness (Catholic schools have higher average achievement than do public schools) and negatively related to the slope (SES effects on math achievement are smaller in Catholic than in public schools). We represent these two hypotheses via two regression equations:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \qquad [2.4a]$$

and

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}, \qquad [2.4b]$$

where

γ_{00} is the mean achievement for public schools;

γ_{01} is the mean achievement difference between Catholic and public schools (i.e., the Catholic school "effectiveness" advantage);

γ_{10} is the average SES-achievement slope in public schools;

γ_{11} is the mean difference in SES-achievement slopes between Catholic and public schools (i.e., the Catholic school "equity" advantage);

u_{0j} is the unique effect of school j on mean achievement holding W_j constant (or conditioning on W_j); and

u_{1j} is the unique effect of school j on the SES-achievement slope holding W_j constant (or conditioning on W_j).

We assume u_{0j} and u_{1j} are random variables with zero means, variances τ_{00} and τ_{11}, respectively, and covariance τ_{01}. Note these variance-covariance components are now *conditional* or *residual* variance-covariance components. That is, they represent the variability in β_{0j} and β_{1j} remaining after controlling for W_j.

It is not possible to estimate the parameters of these regression equations directly, because the outcomes (β_{0j}, β_{1j}) are not observed. However, the data contain information needed for this estimation. This becomes clear if we substitute Equations 2.4a and 2.4b into Equation 2.2, yielding the single prediction equation for the outcome

$$\begin{aligned} Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10}(X_{ij} - \overline{X}_{\cdot j}) + \gamma_{11} W_j (X_{ij} - \overline{X}_{\cdot j}) \\ + u_{0j} + u_{1j}(X_{ij} - \overline{X}_{\cdot j}) + r_{ij}. \end{aligned} \qquad [2.5]$$

Notice that Equation 2.5 is not the typical linear model assumed in standard ordinary least squares (OLS). Efficient estimation and accurate hypothesis testing based on OLS require that the random errors are independent, normally distributed, and have constant variance. In contrast, the random error in Equation 2.5 is of a more complex form, $u_{0j} + u_{1j}(X_{ij} - \overline{X}_{\cdot j}) + r_{ij}$. Such errors are dependent within each school because the components u_{0j} and u_{1j} are common to every student within school j. The errors also have unequal variances, because $u_{0j} + u_{1j}(X_{ij} - \overline{X}_{\cdot j})$ depend on u_{0j} and u_{1j}, which vary across schools, and on the value of $(X_{ij} - \overline{X}_{\cdot j})$, which varies across students. Though standard regression analysis is inappropriate, such models can be estimated by iterative maximum likelihood procedures described in the

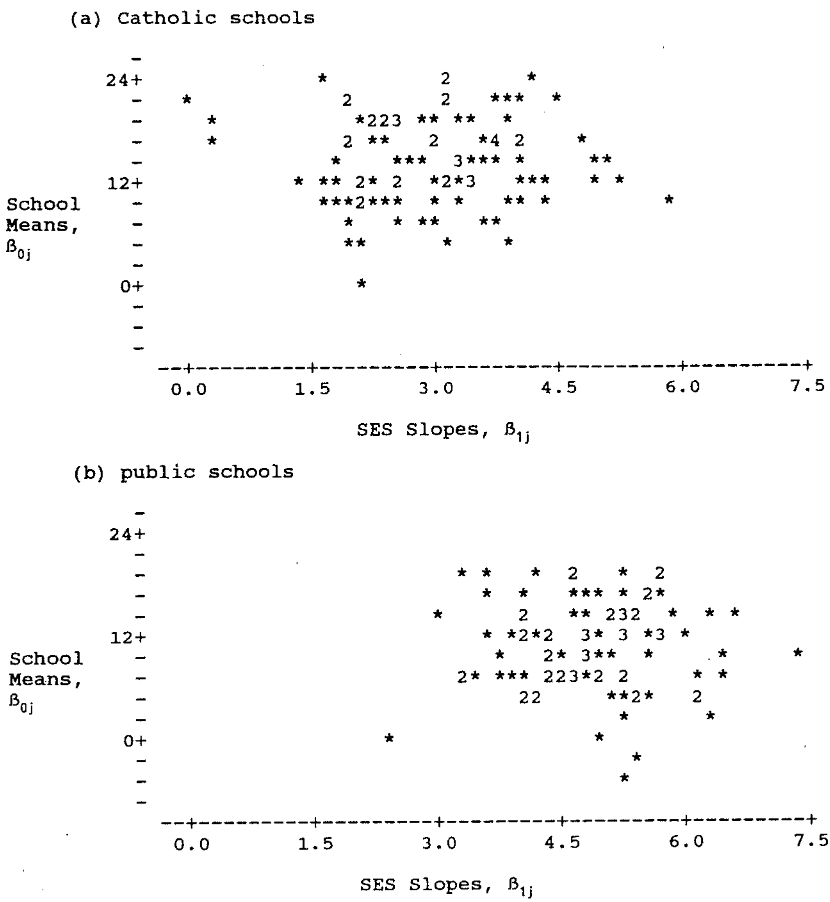

Figure 2.5. Plot of School Means (vertical axis) and SES Slopes (horizontal axis) for 100 Hypothetical Catholic Schools and 100 Hypothetical Public Schools

next chapter. We note that if u_{0j} and u_{1j} were null for every j, Equation 2.5 would be equivalent to an OLS regression model.

Figure 2.5 provides a graphical representation of the model specified in Equation 2.4. Here we see two hypothetical plots of the association between β_{0j} and β_{1j}, one for public and one for Catholic schools. The plots were constructed to reflect Coleman et al.'s (1982) contention that Catholic schools have both higher mean achievement and weaker SES effects than do the public schools.

A General Model and Simpler Submodels

We now generalize our terminology a bit so that it applies to any two-level hierarchical data structure. Equation 2.2 may be labeled the *level-1* model; Equation 2.4 is the *level-2* model, and Equation 2.5 is the *combined* model. In the school-effects application, the level-1 units are students and the level-2 units are schools. The errors r_{ij} are the level-1 random effects and the errors u_{0j} and u_{1j} are level-2 random effects. Moreover, $\mathrm{Var}(r_{ij})$ is the level-1 variance, and $\mathrm{Var}(u_{0j})$, $\mathrm{Var}(u_{1j})$, and $\mathrm{Cov}(u_{0j}, u_{1j})$ are the level-2 variance-covariance components. The β parameters in the level-1 model are level-1 coefficients and the γs are the level-2 coefficients.

Given a single level-1 predictor, X_{ij}, and a single level-2 predictor, W_j, the model given by Equations 2.2, 2.4, and 2.5 is the simplest example of a full hierarchical linear model. When certain sets of terms in this model are set equal to zero, we are left with a set of simpler models, some of which are quite familiar. It is instructive to examine these, both to demonstrate the range of applications of hierarchical linear models and to draw out the connections to more common data analysis methods. The submodels, running from the simpler to the more complex, include the one-way ANOVA model with random effects; a regression model with means-as-outcomes; a one-way analysis of covariance (ANCOVA) model with random effects; a random-coefficients regression model; a model with intercepts- and slopes-as-outcomes; and a model with nonrandomly varying slopes.

One-Way ANOVA with Random Effects

The simplest possible hierarchical linear model is equivalent to a one-way ANOVA with random effects. In this case, β_{1j} in the level-1 model is set to zero for all j, yielding

$$Y_{ij} = \beta_{0j} + r_{ij}. \qquad [2.6]$$

We assume that each level-1 error, r_{ij}, is normally distributed with a mean of zero and a constant level-1 variance, σ^2. Notice that this model predicts the outcome within each level-1 unit with just one level-2 parameter, the intercept, β_{0j}. In this case, β_{0j} is just the mean outcome for the jth unit. That is, $\beta_{0j} = \mu_{Y_j}$.

The level-2 model for the one-way ANOVA with random effects is Equation 2.4a with γ_{01} set to zero:

$$\beta_{0j} = \gamma_{00} + u_{0j}, \qquad [2.7]$$

where γ_{00} represents the grand-mean outcome in the population, and u_{0j} is the random effect associated with unit j and is assumed to have a mean of zero and variance τ_{00}.

Substituting Equation 2.7 into Equation 2.6 yields the combined model

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}, \tag{2.8}$$

which is, indeed, the one-way ANOVA model with grand mean γ_{00}; with a group (level-2) effect, u_{0j}; and with a person (level-1) effect, r_{ij}. It is a random-effects model because the group effects are construed as random. Notice that the variance of the outcome is

$$\mathrm{Var}(Y_{ij}) = \mathrm{Var}(u_{0j} + r_{ij}) = \tau_{00} + \sigma^2. \tag{2.9}$$

Estimating the one-way ANOVA model is often useful as a preliminary step in a hierarchical data analysis. It produces a point estimate and confidence interval for the grand mean, γ_{00}. More important, it provides information about the outcome variability at each of the two levels. The σ^2 parameter represents the within-group variability, and τ_{00} captures the between-group variability. We refer to the hierarchical model of Equations 2.6 and 2.7 as *fully unconditional* in that no predictors are specified at either level 1 or 2.

A useful parameter associated with the one-way random-effects ANOVA is the intraclass correlation coefficient. This coefficient is given by the formula

$$\rho = \tau_{00}/(\tau_{00} + \sigma^2) \tag{2.10}$$

and measures the proportion of the variance in the outcome that is between the level-2 units. See Chapter 4 for an application of the one-way random-effects submodel.

Means-as-Outcomes Regression

Another common statistical problem involves the means from each of many groups as an outcome to be predicted by group characteristics. This submodel consists of Equation 2.6 as the level-1 model and, for the level-2 model,

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}, \tag{2.11}$$

where in this simple case we have one level-2 predictor W_j. Substituting Equation 2.11 into Equation 2.6 yields the combined model:

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + u_{0j} + r_{ij}. \tag{2.12}$$

We note that u_{0j} now has a different meaning as contrasted with that in Equation 2.7. Whereas the random variable u_{0j} had been the deviation of unit j's mean from the grand mean, it now represents the residual

$$u_{0j} = \beta_{0j} - \gamma_{00} - \gamma_{01} W_j.$$

Similarly, the variance in u_{0j}, τ_{00}, is now the residual or conditional variance in β_{0j} after controlling for W_j. The advantages of estimating Equation 2.12 rather than performing a standard regression using sample means-as-outcomes are discussed in Chapter 5.

One-Way ANCOVA with Random Effects

Referring again to the full model (Equations 2.2 and 2.4), let us constrain the level-2 coefficients γ_{01} and γ_{11} and the random effects u_{1j} (for all j) equal to 0. The resulting model would be a one-factor ANCOVA with random effects and a single level-1 predictor as a covariate. The level-1 model is Equation 2.2, but now the predictor X_{ij} is centered around the grand mean. That is,

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{..}) + r_{ij}. \qquad [2.13]$$

The level-2 model becomes

$$\beta_{0j} = \gamma_{00} + u_{0j}, \qquad [2.14a]$$

$$\beta_{1j} = \gamma_{10}. \qquad [2.14b]$$

Notice that the effect of X_{ij} is constrained to be the same fixed value for each level-2 unit as is indicated by Equation 2.14b.

The combined model becomes

$$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \overline{X}_{..}) + u_{0j} + r_{ij}. \qquad [2.15]$$

The only difference between Equation 2.15 and the standard ANCOVA model (cf. Kirk, 1995, chap. 15) is that the group effect here, u_{0j}, is conceived as random rather than fixed. As in ANCOVA, γ_{10} is the pooled within-group regression coefficient of Y_{ij} on X_{ij}. Each β_{0j} is now the mean outcome for each level-2 unit adjusted for differences among these units in X_{ij}. Specifically, $\beta_{0j} = \mu_{Y_j} - \gamma_{10}(\overline{X}_{.j} - \overline{X}_{..})$, where μ_{Y_j} is the mean outcome in school j. We also note that the $\mathrm{Var}(r_{ij}) = \sigma^2$ is now a residual variance after adjusting for the level-1 covariate, X_{ij}.

An extension of the random-effects ANCOVA allows for the introduction of level-2 covariates. For example, if the coefficient γ_{01} is nonnull, the combined model becomes

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_j + \gamma_{10}(X_{ij} - \overline{X}_{..}) + u_{0j} + r_{ij}. \qquad [2.16]$$

This model provides for a level-2 covariate, W_j, while also controlling for the effect of a level-1 covariate, X_{ij}, and the random effects of the level-2 units, u_{0j}. Interestingly, all of the parameters of Equation 2.16 can be estimated using the methods introduced in the next chapter. This is not the case, however, for a classical fixed-effects ANCOVA. Also, the classical ANCOVA model assumes that the covariate effect, γ_{10}, is identical for every group. This homogeneity of regression assumption is easily relaxed using the models described in the next three sections (for randomly varying and nonrandomly varying slopes). We illustrate use of the random-effects ANCOVA model in Chapter 5 in analyzing data on the effectiveness of an instructional innovation on students' writing.

Random-Coefficients Regression Model

All of the submodels discussed above are examples of *random-intercept models*. Only the level-1 intercept coefficient, β_{0j}, was viewed as random. The level-1 slope did not exist in the one-way ANOVA or the means-as-outcomes cases. In the random-effects ANCOVA model, β_{1j} was included but constrained to have a common effect for all groups.

A major class of applications of hierarchical linear models involves studies in which level-1 slopes are conceived as varying randomly over the population of level-2 units. The simplest case of this type is the random-coefficients regression model. In these models, both the level-1 intercept and one or more level-1 slopes vary randomly, but no attempt is made to predict this variation.

Specifically, the level-1 model is identical to Equation 2.2. The level-2 model is still a simplification of Equation 2.4 in that both γ_{01} and γ_{11} are constrained to be null. Hence, the level-2 model becomes

$$\beta_{0j} = \gamma_{00} + u_{0j}, \qquad [2.17a]$$

$$\beta_{1j} = \gamma_{10} + u_{1j}, \qquad [2.17b]$$

where

γ_{00} is the average intercept across the level-2 units;
γ_{10} is the average regression slope across the level-2 units;
u_{0j} is the unique increment to the intercept associated with level-2 unit j; and
u_{1j} is the unique increment to the slope associated with level-2 unit j.

We formally represent the dispersion of the level-2 random effects as a variance-covariance matrix:

$$\mathrm{Var}\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} = \mathbf{T}, \qquad [2.18]$$

where

$\mathrm{Var}(u_{0j}) = \tau_{00}$ = unconditional variance in the level-1 intercepts;
$\mathrm{Var}(u_{1j}) = \tau_{11}$ = unconditional variance in the level-1 slopes; and
$\mathrm{Cov}(u_{0j}, u_{1j}) = \tau_{01}$ = unconditional covariance between the level-1 intercepts and slopes.

Note that we refer to these as unconditional variance-covariance components because no level-2 predictors are included in either Equation 2.17a or 2.17b. Similarly, we refer to Equations 2.17a and 2.17b as an *unconditional* level-2 model.

Substitution of the expressions for β_{0j} and β_{1j} in Equations 2.17a and 2.17b into Equation 2.2 yields a combined model:

$$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \overline{X}_{\cdot j}) + u_{0j} + u_{1j}(X_{ij} - \overline{X}_{\cdot j}) + r_{ij}. \qquad [2.19]$$

This model implies that the outcome Y_{ij} is a function of the average regression equation, $\gamma_{00} + \gamma_{10}(X_{ij} - \overline{X}_{\cdot j})$ plus a random error having three components: u_{0j}, the random effect of unit j on the mean; $u_{1j}(X_{ij} - \overline{X}_{\cdot j})$, where u_{1j} is the random effect of unit j on the slope β_{1j}; and the level-1 error, r_{ij}.

Intercepts- and Slopes-as-Outcomes

The random-coefficients regression model allows us to estimate the variability in the regression coefficients (both intercepts and slopes) across the level-2 units. The next logical step is to model this variability. For example, in Chapter 4, we ask "What characteristics of schools (the level-2 units) help predict why some schools have higher means than others and why some schools have greater SES effects than others?"

Given one level-1 predictor, X_{ij}, and one level-2 predictor, W_j, these questions may be addressed by employing the "full model" of Equations 2.2 and 2.4. Of course, this model may be readily expanded to incorporate the effects of multiple Xs and of multiple Ws (see "Generalizations of the Basic Hierarchical Linear Model").

A Model with Nonrandomly Varying Slopes

In some cases, the analyst will prove quite successful in predicting the variability in the regression slopes, β_{1j}. For example, it might be found that the level-2 predictor W_j in Equation 2.4b does indeed predict the level-1 slope β_{1j}. In fact, the analyst might find that after controlling for W_j the residual variance of β_{1j} (i.e., the variance of the residuals, u_{1j} in Equation 2.4b) is very close to zero. The implication would be that once W_j is controlled, little or no variance in the slopes remains to be explained. For reasons of both statistical efficiency and computational stability (as discussed in Chapter 9), it would be sensible, then, to constrain the values of u_{1j} to be zero. This eliminates τ_{11}, the residual variance of the slope, and τ_{01}, the residual covariance between the slope and the intercept, as parameters to be estimated.

If the residuals u_{1j} in Equation 2.4b are indeed set to zero, the level-2 model for the slopes becomes

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j, \qquad [2.20]$$

and this model, when combined with Equations 2.2 and 2.4a, yields the combined model

$$\begin{aligned} Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10}(X_{ij} - \overline{X}_{\cdot j}) \\ + \gamma_{11} W_j (X_{ij} - \overline{X}_{\cdot j}) + u_{0j} + r_{ij}. \end{aligned} \qquad [2.21]$$

In this model, the slopes do vary from group to group, but their variation is nonrandom. Specifically, as Equation 2.20 shows, the slopes β_{1j} vary strictly as a function of W_j.

We note that Equation 2.21 can be viewed as another example of what we have called a random-intercept model, because β_{0j} is the only component that varies randomly across level-2 units. In general, hierarchical linear models may involve multiple level-1 predictors where any combination of random, nonrandomly varying, and fixed slopes can be specified.

Section Recap

We have been considering a simple hierarchical linear model with a single level-1 predictor, X_{ij}, and a single level-2 predictor, W_j. In this scenario, the level-1 model (Equation 2.2) defines two parameters, the intercept and the slope. At level 2, each of these may be predicted by W_j and each may have a random component of variation, as in Equations 2.4a and 2.4b. The resulting full model, summarized by Equation 2.5, is the most general model we have considered so far. If certain elements of the full model are constrained to be null, we are left with a submodel that may be useful either as preliminary to a full hierarchical analysis or as a more parsimonious summary than the full model.

The six submodels we have considered may be classified in several different ways. We have distinguished between random-intercept models and randomly varying slope models. The one-way random-effects ANOVA model, the means-as-outcomes model, the one-way ANCOVA model, and the model with nonrandomly varying slopes are all random-intercept models. In such models, the variance components are just the level-1 variance, σ^2, and the level-2 variance, τ_{00}. We noted that in the ANOVA and means-as-outcomes models, no level-1 slope exists. In the ANCOVA model, the level-1 slope exists but is constrained or fixed to be invariant across level-2 units. In the nonrandomly varying slope model, slopes were allowed to vary strictly as a function of a known W_j with no additional random component. In contrast, the random-coefficients model and the slopes- and intercepts-as-outcomes models allowed random variation for both the intercepts and slopes.

Another distinction is whether models include *cross-level interaction terms* such as $\gamma_{11} W_j(X_{ij} - \overline{X}_{\cdot j})$. In general, the combined model will include such cross-level interaction terms whenever we seek to predict variation in a slope. Such terms appear in two of our submodels: the intercepts- and slopes-as-outcomes model and the nonrandomly varying slope model.

Generalizations of the Basic Hierarchical Linear Model

Multiple *X*s and Multiple *W*s

Suppose now that the analyst wishes to use information about a second level-1 predictor. Let X_{1ij} denote the original X discussed above and let X_{2ij} denote the second level-1 predictor. For now, assume that there is still just a single level-2 predictor, W_j. The level-1 model, assuming group-mean centering for both X_{1ij} and X_{2ij}, becomes

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{1ij} - \overline{X}_{1\cdot j}) + \beta_{2j}(X_{2ij} - \overline{X}_{2\cdot j}) + r_{ij}. \qquad [2.22]$$

Again, we have three options for modeling β_{2j}. One option is that the effect of X_{2ij} is constrained to be invariant across level-2 units, implying

$$\beta_{2j} = \gamma_{20},$$

where γ_{20} is the common effect of X_{2ij} in every level-2 unit. We say that the effect of β_{2j} is *fixed* across level-2 units.

A second option would be to model the slope β_{2j} as a function of an average value, γ_{20}, plus a random effect associated with each level-2 unit:

$$\beta_{2j} = \gamma_{20} + u_{2j}. \qquad [2.23]$$

Here β_{2j} is *random*. Notice that Equation 2.23 specifies no predictors for β_{2j}. Suppose, however, that this slope depends on W_j. One might then formulate the slopes-as-outcomes model:

$$\beta_{2j} = \gamma_{20} + \gamma_{21} W_j + u_{2j}. \qquad [2.24]$$

According to this model, part of the variation of the slope β_{2j} can be predicted by W_j, but a random component, u_{2j}, remains unexplained. On the other hand, it may be that once the effect of W_j is taken into account, the residual variation in β_{2j}—that is, $\text{Var}(u_{2j}) = \tau_{22}$—is negligible. Then a model constraining that residual variation to be null would be sensible:

$$\beta_{2j} = \gamma_{20} + \gamma_{21} W_j. \qquad [2.25]$$

In this third case, β_{2j} is a *nonrandomly varying* slope because it varies strictly as a function of the predictor W_j.

So far we have been interested in just a single level-2 predictor, W_j. The introduction of multiple W_js is straightforward. Further, the level-2 model does not need to be identical for each equation. One set of W_js may apply for the intercept, a different set be used for β_{1j}, another set for β_{2j}, and so on. When nonparallel specification is employed, however, extra care must be exercised in the interpretation of the results (see Chapter 9).

Generalization of the Error Structures at Level 1 and Level 2

The model specified in Equations 2.2 and 2.4 assumes homogeneous errors at both level 1 and level 2. This assumption is quite acceptable for a broad class of multilevel problems. Most published applications have been based on this assumption, as are most of the examples discussed in Chapter 5 through 8.

The model can easily be extended, however, to more complex error structures at both levels. The level-1 variance might be different for each level-2 unit and denoted σ_j^2, or it might be a function of some measured level-1 characteristic. (The modeling framework for this extension appears in Chapter 5.) Similarly, at level 2, a different covariance structure might exist for distinct subsets of level-2 units. This would result in different T matrices estimated for different subsets of level-2 units.

Extensions Beyond the Basic Two-Level Hierarchical Linear Model

The core ideas introduced in this chapter in the context of two-level models extend directly to models with three or more levels. These extensions are described and illustrated in Chapter 8. A common feature of the basic hierarchical linear model, regardless of the number of levels, is that the outcome variable at level 1, Y, is continuous and assumed normally distributed, conditional on the level-1 predictors included in the model. Over the last decade, extensions beyond the basic hierarchical linear model framework have been advanced to include dichotomous level-1 outcomes, count data, and categorical outcomes. Models for missing data, latent variable effects, and more complex data designs, including crossed random effects, have also appeared. Although the estimation methods are more complex for these extensions, the basic conceptual ideas and modeling framework extend quite naturally. In general, the range of modeling possibilities is now much richer than when we authored the first edition of this book. Part III, which is new to the second edition, introduces these new developments.

Choosing the Location of X and W (*Centering*)

In all quantitative research, it is essential that the variables under study have precise meaning so that statistical results can be related to the theoretical concerns that motivate the research. In the case of hierarchical linear models, the intercept and slopes in the level-1 model become outcome variables at level 2. It is vital that the meaning of these outcome variables be clearly understood.

The meaning of the intercept in the level-1 model depends on the location of the level-1 predictor variables, the Xs. We know, for example, that in the simple model

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}, \qquad [2.26]$$

the intercept, β_{0j}, is defined as the expected outcome for a student attending school j who has a value of zero on X_{ij}. If the researcher is to make sense of models that account for variation in β_{0j}, the choice of a metric for all level-1 predictors becomes important. In particular, if an X_{ij} value of zero is not meaningful, then the researcher may want to transform X_{ij}, or "choose a location for X_{ij}" that will render β_{0j} more meaningful. In some cases, a proper choice of location will be required in order to ensure numerical stability in estimating hierarchical linear models.

Similarly, interpretations regarding the intercepts in the level-2 models (i.e., γ_{00} and γ_{10} in Equations 2.4a and 2.4b) depend on the location of the W_j variables. The numerical stability of estimation is not affected by the location for the Ws, but a suitable choice will ease interpretation of results. We describe below some common choices for the location of the Xs and Ws.

Location of the *X*s

We consider four possibilities for the location of X: the natural X metric, centering around the grand mean, centering around the group mean, and other locations for X. We assume that X is measured on an interval scale. The case of dummy variables is considered separately.

The Natural X *Metric.* Although the natural X metric may be quite appropriate in some applications, in others this may lead to nonsensical results. For example, suppose X is a score on the Scholastic Aptitude Test (SAT), which ranges from 200 to 800. Then the intercept, β_{0j}, will be the expected outcome for a student in school j who had an SAT of zero. The β_{0j} parameter is meaningless in this instance because the minimum score on the test is 200. In such cases, the correlation between the intercept and slope will tend toward -1.0. As a result, the intercept is essentially determined by the slope. Schools with strong positive SAT-outcome slopes will tend to have very low intercepts. In contrast, schools where the SAT slope is negligible will tend to have much higher intercepts.

In some applications, of course, an X value of zero will in fact be meaningful. For example, if X is the dosage of an experimental drug, $X_{ij} = 0$ implies that subject i in group j had no exposure to the drug. As a result, the intercept β_{0j} is the expected outcome for such a subject. That is, $\beta_{0j} = \mathrm{E}(Y_{ij}|X_{ij} = 0)$. We wish to emphasize that it is always important to consider the meaning of $X_{ij} = 0$ because it determines the interpretation of β_{0j}.

Centering Around the Grand Mean. It is often useful to center the variable X around the grand mean, as discussed earlier (see "One-Way ANCOVA with Random Effects"). In this case, the level-1 predictors are of the form

$$(X_{ij} - \overline{X}_{..}). \quad [2.27]$$

Now, the intercept, β_{0j}, is the expected outcome for a subject whose value on X_{ij} is equal to the grand mean, $\overline{X}_{..}$. This is the standard choice of location for X_{ij} in the classical ANCOVA model. As is the case in ANCOVA, grand-mean centering yields an intercept that can be interpreted as an adjusted mean for group j,

$$\beta_{0j} = \mu_{Y_j} - \beta_{1j}(\overline{X}_{\cdot j} - \overline{X}_{..}).$$

Similarly, the $\mathrm{Var}(\beta_{0j}) = \tau_{00}$ is the variance among the level-2 units in the adjusted means.

Centering Around the Level-2 Mean (Group-Mean Centering). Another option is to center the original predictors around their corresponding level-2 unit means:

$$(X_{ij} - \overline{X}_{\cdot j}). \quad [2.28]$$

In this case, the intercept β_{0j} becomes the unadjusted mean for group j. That is,

$$\beta_{0j} = \mu_{Y_j} \quad [2.29]$$

and $\mathrm{Var}(\beta_{0j})$ is now just the variance among the level-2 unit means, μ_{Y_j}.

Other Locations for X. Specialized choices of location for X are often sensible. In some cases, the population mean for a predictor may be known and the investigator may wish to define the intercept β_{0j} as the expected outcome in group j for the "average person in the population." In this case, the level-1 predictor would be the original value of X_{ij} minus the population mean.

In applications of two-level hierarchical linear models to the study of growth, the data involve time-series observations so that the level-1 units are occasions and the level-2 units are persons. The investigator may wish to define the metric of the level-1 predictors such that the intercept is the expected outcome for person i at a specific time point of theoretical interest (e.g., entry to school). So long as the data encompass this time point, such a definition is quite appropriate. Examples of this sort are illustrated in Chapters 6 and 8.

Dummy Variables. Consider the familiar level-1 model

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}, \qquad [2.30]$$

where X_{ij} is now an indicator or dummy variable. Suppose, for example, that X_{ij} takes on a value of 1 if subject i in school j is a female and 0 if not. In this case, the intercept β_{0j} is defined as the expected outcome for a male student in group j (i.e., the predicted value for student with $X_{ij} = 0$). We note in this case that $\text{Var}(\beta_{0j}) = \tau_{00}$ will be the variance in the male outcome means across schools.

Although it may seem strange at first to center a level-1 dummy variable, this is appropriate and often quite useful. Suppose, for example, that the indicator variable for sex is centered around the grand mean, $\overline{X}_{..}$. This centered predictor can take on two values. If the subject is female, $X_{ij} - \overline{X}_{..}$ will equal the proportion of male students in the sample. If the subject is male, $X_{ij} - \overline{X}_{..}$ will be equal to minus the proportion of female students. As in the case of continuous level-1 predictors centered around the respective grand means, the intercept, β_{0j}, is the adjusted mean outcome in unit j. In this case, it is adjusted for differences among units in the percentage of female students.

Alternatively, we might use group-mean centering. For females, $X_{ij} - \overline{X}_{\cdot j}$ will take on the value equal to the proportion of male students in school j; for males, $X_{ij} - \overline{X}_{\cdot j}$ will take on a value equal to minus the proportion of female students in school j. The fact that X_{ij} is a dummy variable does not change the interpretation given to β_{0j} when group-mean centering is employed. The intercept still represents the average outcome for unit j, μ_{Y_j}.

In sum, several locations of dichotomous predictors will produce meaningful intercepts. Again, it is incumbent on the researcher to take this location into account in interpreting results. Care is especially needed when there are multiple dummy variables. For example, in a school-effects study with indicators for whites, females, and students with preprimary education, the intercept for school j might be the expected outcome for a non-white male student with no preprimary experience. This may or may not be the intercept the investigator wants. Again we offer the general caveat—be conscious of the choice of location for each level-1 predictor because it has implications for interpretation of β_{0j}, $\text{Var}(\beta_{0j})$, and by implication, all of the covariances involving β_{0j}.

In general, sensible choices of location depend on the purposes of the research. No single rule covers all cases. It is important, however, that the researcher carefully consider choices of location in light of those purposes; and it is vital to keep the location in mind while interpreting results.

In addition, the choice of location for the level-1 predictors can, under certain circumstances, also influence the estimation of the level-2 variance-covariance components, **T**, and random level-1 coefficients, β_{qj}. Complications can occur in the context of both organizational research and growth curve applications. The reader is referred to Chapters 5 and 6, respectively, for a further discussion of these technical considerations.

Location of *W*s

In general, the choice of location for the Ws is not as critical as for the level-1 predictors. Problems of numerical instability are less likely, except when cross-product terms are introduced at level 2 (e.g., a predictor set of the form W_{1j}, W_{2j}, and $W_{1j}W_{2j}$). All of the γ coefficients can be easily interpreted whatever choice of metric (or nonchoice) is made for level-2 predictors. Nevertheless, it is often convenient to center all of the level-2 predictors around their corresponding grand means, for example, $W_{1j} - \overline{W}_{1}.$.

Summary of Terms and Notation Introduced in This Chapter

A Simple Two-Level Model

Hierarchical form:

$$\begin{array}{ll} \text{Level 1 (e.g., students)} & Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}, \\ \text{Level 2 (e.g., schools)} & \beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}, \\ & \beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}. \end{array}$$

Model in *combined form:*

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_j + \gamma_{11}X_{ij}W_j + u_{0j} + u_{1j}X_{ij} + r_{ij},$$

where we assume:

$$\mathrm{E}(r_{ij}) = 0, \qquad \mathrm{Var}(r_{ij}) = \sigma^2,$$

$$\mathrm{E}\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \qquad \mathrm{Var}\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} = \mathbf{T},$$

$$\mathrm{Cov}(u_{0j}, r_{ij}) = \mathrm{Cov}(u_{1j}, r_{ij}) = 0.$$

Notation and Terminology Summary

There are $i = 1, \ldots, n_j$ level-1 units nested with $j = 1, \ldots, J$ level-2 units. We speak of student i nested within school j.

β_{0j}, β_{1j} are level-1 coefficients. These can be of three forms:

fixed level-1 coefficients (e.g., β_{1j} in the one-way random-effects ANCOVA model, Equation 2.14b)

nonrandomly varying level-1 coefficients (e.g., β_{1j} in the nonrandomly-varying-slopes model, Equation 2.20)

random level-1 coefficients (e.g., β_{0j} and β_{1j} in the random-coefficient regression model [Equations 2.17a and 2.17b] and in the intercepts- and slopes-as-outcomes model [Equations 2.4a and 2.4b])

$\gamma_{00}, \ldots, \gamma_{11}$ are level-2 coefficients and are also called fixed effects.

X_{ij} is a level-1 predictor (e.g., student social class, race, and ability).

W_j is a level-2 predictor (e.g., school size, sector, social composition).

r_{ij} is a level-1 random effect.

u_{0j}, u_{1j} are level-2 random effects.

σ^2 is the level-1 variance.

$\tau_{00}, \tau_{01}, \tau_{11}$ are level-2 variance-covariance components.

Some Definitions

Intraclass correlation coefficient (see "One-Way ANOVA with Random Effects"):

$$\rho = \tau_{00}/(\sigma^2 + \tau_{00}).$$

This coefficient measures the proportion of variance in the outcome that is between groups (i.e., the level-2 units). It is also sometimes called the *cluster effect.* It applies only to random-intercept models (i.e., $\tau_{11} = 0$).

Unconditional variance-covariance of β_{0j}, β_{1j} are the values of the level-2 variances and covariances based on the random-coefficient regression model.

Conditional or residual variance-covariance of β_{0j}, β_{1j} are the values of the level-2 variances and covariances after level-2 predictors have been added for β_{0j} and β_{1j} (see, e.g., Equations 2.4a and 2.4b).

Submodel Types

One-way random-effects ANOVA model involves no level-1 or level-2 predictors. We call this a *fully unconditional* model.

Random-intercept model has only one random level-1 coefficient, β_{0j}.

Means-as-outcomes regression model is one form of a random-intercept model.

One-way random-effects ANCOVA model is a classic ANCOVA model, except that the level-2 effects are viewed as random.

Random-coefficients regression model allows all level-1 coefficients to vary randomly. This model is *unconditional at level 2*.

Centering Definitions	**Implications for β_{0j}**
X_{ij} in the natural metric	$\beta_{0j} = \mathrm{E}(Y_{ij} \mid X_{ij} = 0)$
$(X_{ij} - \overline{X}_{..})$ called grand-mean centering	$\beta_{0j} = \mu_{Y_j} - \beta_{1j}(X_{ij} - \overline{X}_{..})$ (i.e., adjusted level-2 means)
$(X_{ij} - \overline{X}_{.j})$ called group-mean centering	$\beta_{0j} = \mu_{Y_j}$ (i.e., level-2 means)
X_{ij} centered at some theoretically chosen location for X	$\beta_{0j} = \mathrm{E}(Y_{ij} \mid X_{ij} =$ chosen centering location for $X)$