MULTI-ITEM MEASURES

The item is the basic unit of a psychological scale. It is a statement or question in clear, unequivocal terms about the measured characteristics (Haladyna, 1994). In social science, scales are used to assess people’s social characteristics, such as attitudes, personal values, opinions, emotional states, personality needs, and description of their living environment.

An item is a “mini” measure that has a molecular score (Thornlakke, 1967). When used in social science, multi-item measures can be superior to a single, straightforward question. There are two reasons. First, the RELIABILITY of a multi-item measure is higher than a single question. With a single question, people are less likely to give consistent answers over time. Many things can influence people’s response (e.g., mood, specific thing they encountered that day). They may choose yes to a question one day and no the other day. It is also possible that people give a wrong answer or interpret the question differently over time.

On the other hand, a multi-item measure has several questions targeting the same social issue, and the final composite score is based on all questions. People are less likely to make the above mistakes to multiple items, and the composite score is more consistent over time. Thus, the multi-item measure is more reliable than a single question. Second, the VALIDITY of a multi-item measure can be higher than a single question. Many measured social characteristics broad in scope and simply cannot be assessed with a single question. Multi-item measures will be necessary to cover more content of the measured characteristic and to fully and completely reflect the construct domain.

These issues are best illustrated with an example. To assess people’s job satisfaction, a single-item measure could be as follows:

I’m not satisfied with my work. (1 = disagree, 2 = slightly disagree, 3 = uncertain, 4 = slightly agree, 5 = agree)

To this single question, people’s responses can be inconsistent over time. Depending on their mood or specific things they encountered at work that day, they might respond very differently to this single question. Also, people may make mistakes when reading or responding. For example, they might not notice the word not and agree when they really disagree.

Thus, this single-item measure about job satisfaction can be notoriously unreliable. Another problem is that people’s feelings toward their jobs may not be simple. Job satisfaction is a very broad issue, and it includes many aspects (e.g., satisfaction with the supervisor, satisfaction with coworkers, satisfaction with work content, satisfaction with pay, etc.). Subjects may like certain aspects of their jobs but not others. The single-item measure will oversimplify people’s feelings toward their jobs.

A multi-item measure can reduce the above problems. The results from a multi-item measure should be more consistent over time. With multiple items, random errors could average out (Spector, 1992). That is, with 20 items, if a respondent makes an error on 1 item, the impact on the overall score is quite minimal. More important, a multi-item measure will allow subjects to describe their feelings about different aspects of their jobs. This will greatly improve the precision and validity of the measure. Therefore, multi-item measures are one of the most important and frequently used tools in social science.

—Cong Liu

See also SCALING

REFERENCES


MULTILEVEL ANALYSIS

Most of statistical inference is based on replicated observations of UNITS OF ANALYSIS of one type (e.g., a sample of individuals, countries, or schools). The analysis of such observations usually is based on the
assumption that either the sampled units themselves or
the corresponding RESIDUALS in some statistical model
are independent and identically distributed. However,
the complexity of social reality and social science
theories often calls for more COMPLEX DATA SETS,
which include units of analysis of more than one
type. Examples are studies on educational achieve-
ment, in which pupils, teachers, classrooms, and
schools might all be important units of analysis; organi-
zational studies, with employees, departments, and
firms as units of analysis; cross-national comparative
research, with individuals and countries (perhaps also
regions) as units of analysis; studies in GENERALIZ-
ABILITY THEORY, in which each factor defines a type
of unit of analysis; and META-ANALYSIS, in which the
collected research studies, the research groups that pro-
duced them, and the subjects or respondents in these
studies are units of analysis and sources of unexplained
variation. Frequently, but by no means always, units
of analysis of different types are hierarchically nested
(e.g., pupils are nested in classrooms, which, in turn,
are nested in schools). Multilevel analysis is a general
term referring to statistical methods appropriate for
the analysis of data sets comprising several types of
unit of analysis. The levels in the multilevel analysis
are another name for the different types of unit of
analysis. Each level of analysis will correspond to a
POPULATION, so that multilevel studies will refer to
several populations—in the first example, there are
four populations: of pupils, teachers, classrooms, and
schools. In a strictly nested data structure, the most
detailed level is called the first, or the lowest, level. For
example, in a data set with pupils nested in classrooms
nested in schools, the pupils constitute Level 1, the
classrooms Level 2, and the schools Level 3.

HIERARCHICAL LINEAR MODEL

The most important methods of multilevel analysis
are variants of REGRESSION analysis designed for
hierarchically nested data sets. The main model is the
HIERARCHICAL LINEAR MODEL (HLM), an extension of
the GENERAL LINEAR MODEL in which the probabil-
ity model for the errors, or residuals, has a structure
reflecting the hierarchical structure of the data. For
this reason, multilevel analysis often is called hierar-
chical linear modeling. As an example, suppose that
a researcher is studying how annual earnings of
college graduates well after graduation depend on aca-
demic achievement in college. Let us assume that
the researcher collected data for a reasonable number
of colleges—say, more than 30 colleges that can be
regarded as a sample from a specific population of
colleges, with this population being further specified
to one or a few college programs—and, for each of
these colleges, a random sample of the students who
graduated 15 years ago. For each student, information
was collected on the current income (variable Y)
and the grade point average in college (denoted by the
variable X in a metric where X = 0 is the minimum
passing grade). Graduates are denoted by the letter i
and colleges by j. Because graduates are nested in col-
lleges, the numbering of graduates i may start from 1 for
each college separately, and the variables are denoted
by Y_{ij} and X_{ij}. The analysis could be based for college
j on the model

$$Y_{ij} = a_j + b_j X_{ij} + E_{ij}.$$ 

This is just a linear regression model, in which the
INTERCEPTS a_j and the regression coefficients b_j
depend on the college and therefore are indicated with
the subscript j. The fact that colleges are regarded as
a random sample from a population is reflected by the
assumption of random variation for the intercepts a_j
and regression coefficients b_j. Denote the population
mean (in the population of all colleges) of the inter-
cepts by \(\bar{a}\) and the college-specific deviations by \(U_{0j}\),
so that \(a_j = a + U_{0j}\). Similarly, split the regression
coefficients into the population mean and the college-
specific deviations \(b_j = b + U_{1j}\). Substitution of these
equations then yields

$$Y_{ij} = a + b X_{ij} + U_{0j} + U_{1j} X_{ij} + E_{ij}.$$ 

This model has three different types of residuals: the
so-called Level-1 residual \(E_{ij}\) and the Level-2 resid-
uals \(U_{0j}\) and \(U_{1j}\). The Level-1 residual varies over
the population of graduates; the Level-2 residuals vary
over the population of colleges. The residuals can be
interpreted as follows. For colleges with a high value
of \(U_{0j}\), their graduates with the minimum passing
grade \(X = 0\) have a relatively high expected income—
namely, \(a + U_{0j}\). For colleges with a high value of \(U_{1j}\),
the effect of one unit GPA extra on the expected income
of their graduates is relatively high—namely, \(b + U_{1j}\).
Graduates with a high value of \(E_{ij}\) have an income
that is relatively high, given their college \(j\) and their
GPA \(X_{ij}\).

This equation is an example of the HLM; in its
general form, this model can have more than one
independent variable. The first part of the equation, \( a + bX_{ij} \), is called the fixed part of the model; this is a linear function of the independent variables, just like in linear regression analysis. The second part, \( U_{0ij} + U_{1ij}X_{ij} + E_{ij} \), is called the random part and is more complicated than the random residual in linear regression analysis, as it reflects the unexplained variation \( E_{ij} \) between the graduates as well as the unexplained variation \( U_{0ij} + U_{1ij}X_{ij} \) between the colleges. The random part of the model is what distinguishes the hierarchical linear model from the general linear model. The simplest nontrivial specification for the random part of a two-level model is a model in which only the intercept varies between Level-2 units, but the regression coefficients are the same across Level-2 units. This is called the random intercept model, and for our example, it reads

\[
Y_{ij} = a + bX_{ij} + U_{0ij} + E_{ij}.
\]

Models in which also the regression coefficients vary randomly between Level-2 units are called random slope models (referring to graphs of the regression lines, in which the regression coefficients are the slopes of the regression lines).

The dependent variable \( Y \) in the HLM always is a variable defined at the lowest (i.e., most detailed) level of the hierarchy. An important feature of the HLM is that the independent, or explanatory, variables can be defined at any of the levels of analysis. In the example of the study of income of college graduates, suppose that the researcher is interested in the effect on earnings of alumni of college quality, as measured by college rankings, and that some meaningful college ranking score \( Z_j \) is available. In the earlier model, the college-level residuals \( U_{0ij} \) and \( U_{1ij} \) reflect unexplained variability between colleges. This variability could be explained partially by the college-level variable \( Z_j \), according to the equations

\[
a_j = a + c_0Z_j + U_{0j}, \quad b_j = b + c_1Z_j + U_{1j},
\]

which can be regarded as linear regression equations at Level 2 for the quantities \( a_j \) and \( b_j \), which are themselves not directly observable. Substitution of these equations into the Level-1 equation \( Y_{ij} = a_j + b_jX_{ij} + E_{ij} \) yields the new model,

\[
Y_{ij} = a + bX_{ij} + c_0Z_j + c_1X_{ij}Z_j + U_{0j} + U_{1ij}X_{ij} + E_{ij},
\]

where the parameters \( a \) and \( b \) and the residuals \( E_{ij}, U_{0ij}, \) and \( U_{1ij} \) now have different meanings than in the earlier model. The fixed part of this model is extended compared to the earlier model, but the random part has retained the same structure. The term \( c_1X_{ij}Z_j \) in the fixed part is the interaction effect between the Level-1 variable \( X \) and the Level-2 variable \( Z \). The regression coefficient \( c_1 \) expresses how much the college context \( Z \) modifies the effect of the individual achievement \( X \) on later income \( Y \); such an effect is called a cross-level interaction effect. The possibility of expressing how context (the "macro level") affects relations between individual-level variables (the "micro level") is an important reason for the popularity of multilevel modeling (see DiPrete & Forristal, 1994).

A parameter that describes the relative importance of the two levels in such a data set is the intraclass correlation coefficient, described in the entry with this name and also in the entry on variance component models. The similar variance ratio, when applied to residual (i.e., unexplained) variances, is called the residual intraclass correlation coefficient.

**ASSUMPTIONS, ESTIMATION, AND TESTING**

The standard assumptions for the HLM are the linear model expressed by the model equation, normal distributions for all residuals, and independence of the residuals for different levels and for different units in the same level. However, different residuals for the same unit, such as the random intercept \( U_{0ij} \) and the random slope \( U_{1ij} \) in the model above, are allowed to be correlated; they are assumed to have a multivariate normal distribution. With these assumptions, the HLM for the example above implies that outcomes for graduates of the same college are correlated due to the influences from the college—technically, due to the fact that their equations for \( Y_{ij} \) contain the same college-level residuals \( U_{0ij} \) and \( U_{1ij} \). This dependence between different cases is an important departure from the assumptions of the more traditional general linear model used in regression analysis.

The parameters of the HLM can be estimated by the maximum likelihood method. Various algorithms have been developed mainly in the 1980s (cf. Goldstein, 2003; Longford, 1993); one important algorithm is an iterative reweighted least squares algorithm (see the entry on generalized least squares), which alternates between estimating the regression coefficients in the fixed part and the parameters of
the random part. The regression coefficients can be tested by \( t \)-tests or Wald tests. The parameters defining the structure of the random part can be tested by likelihood ratio tests (also called deviance tests) or by chi-squared tests. These methods have been made available since the 1980s in dedicated multilevel software, such as HLM and MLwiN, and later also in packages that include multilevel analysis among a more general array of methods, such as M-Plus, and in some general statistical packages, such as SAS and SPSS. An overview of software capabilities is given in Goldstein (2003).

**MULTIPLE LEVELS**

As was illustrated already in the examples, it is not uncommon that a practical investigation involves more than two levels of analysis. In educational research, the largest contributions to achievement outcomes usually are determined by the pupil and the teacher, but the social context provided by the group of pupils in the classroom and the organizational context provided by the school, as well as the social context defined by the neighborhood, may also have important influences. In a study of academic achievement of pupils, variables defined at each of these levels of analysis could be included as explanatory variables. If there is an influence of some level of analysis, then it is to be expected that this influence will not be completely captured by the variables measured for this level of analysis, but there will be some amount of unexplained variation between the units of analysis for this level. This should then be reflected by including this unexplained variation as random residual variability in the model. The first type of residual variability is the random main effect of the units at this level, exemplified by the random intercepts \( U_{ij} \) in the two-level model above. In addition, it is possible that the effects of numerical variables (such as pupil-level variables) differ across the units of the level under consideration, which can again be modeled by random slopes such as the \( U_{ij} \) above. An important type of conclusion of analyses with multiple levels of analysis is the partitioning of unexplained variability over the various levels. This is discussed for models without random slopes in the entry on VARIANCE COMPONENT MODELS. How much unexplained variability is associated with each of the levels can provide the researcher with important directions about where to look for further explanation.

Levels of analysis can be nested or crossed. One level of analysis—the lower level—is said to be nested in another, higher level if the units of the higher level correspond to a partition into subsets of the units of the lower level (i.e., each unit of the lower level is contained in exactly one unit of the higher level). Otherwise, the levels are said to be crossed. Crossed levels of analysis often are more likely to lead to difficulties in the analysis than nested levels. Estimation algorithms may have more convergence problems, the empirical conclusions about partitioning variability over the various levels may be less clear-cut, and there may be more ambiguity in conceptual and theoretical modeling.

The use of models with multiple levels of analysis requires a sufficiently rich data set on which to base the statistical analysis. Note that for each level of analysis, the units in the data set constitute a sample from the corresponding population. Although any rules of thumb should be taken with a grain of sand, a sample size less than 20 (i.e., a level of analysis represented by less than 20 units) usually will give only quite restricted information about this population (i.e., this level of analysis), and sample sizes less than 10 should be regarded with suspicion.

**LONGITUDINAL DATA**

In longitudinal research, the HLM also can be used fruitfully. In the most simple longitudinal data structure, with repeated measures on individuals, the repeated measures constitute the lower (first) and the individuals the higher (second) levels. Mostly, there will be a meaningful numerical time variable: For example, in an experimental study, this may be the time since onset of the experimental situation, and in a developmental study, this may be age. Especially for nonbalanced longitudinal data structures, in which the numbers and times of observations differ between individuals, multilevel modeling may be a natural and very convenient method. The dependence of the outcome variable on the time dimension is a crucial aspect of the model. Often, a linear dependence is a useful first approximation. This amounts to including the time of measurement as an explanatory variable; a random slope for this variable represents differential change (or growth) rates for different individuals. Often, however, dependence on time is nonlinear. In some cases, it will be possible to model this while remaining within the HLM by using several nonlinear transformations.
(e.g., polynomials or splines) of time and postulating a model that is linear in these transformed time variables (see Snijders & Bosker, 1999, chap. 12). In other cases, it is better to forgo the relative simplicity of linear modeling and construct models that are not linear in the original or transformed variables or for which the Level-1 residuals are autocorrelated (cf. Verbeke & Molenberghs, 2000).

NONLINEAR MODELS

The assumption of normal distributions for the residuals is not always appropriate, although sometimes this assumption can be made more realistic by transformations of the dependent variable. In particular, for dichotomous or discrete dependent variables, other models are required. Just as the GENERALIZED LINEAR MODEL is an extension of the general linear model of regression analysis, nonlinear versions of the HLM also provide the basis of, for example, multilevel versions of LOGISTIC REGRESSION and LOGIT MODELS. These are called hierarchical generalized linear models or generalized linear mixed models (see the entry on HIERARCHICAL NONLINEAR MODELS).

—Tom A. B. Snijders

REFERENCES


MULTIMETHOD-MULTITRAIT RESEARCH. See MULTIMETHOD RESEARCH

MULTIMETHOD RESEARCH

Multimethod research entails the application of two or more sources of data or research methods to the investigation of a research question or to different but highly linked research questions. Such research is also frequently referred to as mixed methodology. The rationale for mixed-method research is that most social research is based on findings deriving from a single research method and, as such, is vulnerable to the accusation that any findings deriving from such a study may lead to incorrect inferences and conclusions if MEASUREMENT ERROR is affecting those findings. It is rarely possible to estimate how much measurement error is having an impact on a set of findings, so that monomethod research is always suspect in this regard.

MIXED-METHOD RESEARCH AND MEASUREMENT

The rationale of mixed-method research is underpinned by the principle of TRIANGULATION, which implies that researchers should seek to ensure that they are not overreliant on a single research method and should instead employ more than one measurement procedure when investigating a research problem. Thus, the argument for mixed-method research, which in large part accounts for its growth in popularity, is that it enhances confidence in findings.

In the context of measurement considerations, mixed-method research might be envisioned in relation to different kinds of situations. One form might be that when one or more constructs that are the focus of an investigation have attracted different measurement efforts (such as different ways of measuring levels of job satisfaction), two or more approaches to measurement might be employed in combination. A second form might entail employing two or more methods of data collection. For example, in developing an approach to the examination of the nature of jobs in a firm, we might employ structured observation and structured interviews concerning apparently identical
Hierarchical Linear Models
Applications and Data Analysis Methods
Second Edition

Stephen W. Raudenbush
Anthony S. Bryk

Advanced Quantitative Techniques
in the Social Sciences Series

Sage Publications
International Educational and Professional Publisher
Thousand Oaks • London • New Delhi
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 $\beta_0$ and slope $\beta_1$. Thus, the regression equation for the data is

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

The intercept, $\beta_0$, is defined as the expected math achievement of a student whose SES is zero. The slope, $\beta_1$, is the expected change in math achievement associated with a unit increase in SES. The error term, $r_i$, represents a unique effect associated with person $i$. Typically, we assume that $r_i$ is normally distributed with a mean of zero and variance $\sigma^2$, that is, $r_i \sim N(0, \sigma^2)$.

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: $X_i - \bar{X}$, where $\bar{X}$ is the mean SES in the school. If we now plot $Y_i$ as a function of $X_i - \bar{X}$ (see Figure 2.2) with the
The Logic of Hierarchical Linear Models

A 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

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_j) + r_{ij},$$

where for simplicity we assume that $r_{ij}$ is normally distributed with homogeneous variance across schools, that is, $r_{ij} \sim N(0, \sigma^2)$. 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 $(\beta_{0j}, \beta_{1j})$. It is often sensible and convenient to assume that the intercept and slope have a bivariate normal distribution across the population of schools. Let

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

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

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

where

$\gamma_0$ is the average school mean for the population of schools;

$\tau_{00}$ is the population variance among the school means;

$\gamma_1$ is the average SES-achievement slope for the population;

$\tau_{11}$ is the population variance among the slopes; and

$\tau_{01}$ is the population covariance between slopes and intercepts.

A positive value of $\tau_{01}$ 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:

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

In reality, we rarely know the true values of the population parameters we have introduced ($\gamma_0, \gamma_1, \tau_{00}, \tau_{01}$) nor of the true individual school means and slopes $(\beta_{0j} \text{ and } \beta_{1j})$. 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 $\beta_{0j}$ and
\[ \beta_{0j} = \gamma_0 \omega_j + u_{0j} \]  
and  
\[ \beta_{1j} = \gamma_1 \omega_j + u_{1j}, \]  

where

\[ \gamma_0 \] is the mean achievement for public schools;

\[ \gamma_1 \] is the mean achievement difference between Catholic and public schools (i.e., the Catholic school "effectiveness" advantage);

\[ \gamma_2 \] is the average SES-achievement slope in public schools;

\[ \gamma_3 \] 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 \( \omega_j \) constant (or conditioning on \( \omega_j \)); and

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

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

It is not possible to estimate the parameters of these regression equations directly, because the outcomes \( (\beta_{0j}, \beta_{1j}) \) 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

\[ Y_{ij} = \gamma_0 + \gamma_1 (X_{ij} - \bar{X}_j) + \gamma_2 W_j (X_{ij} - \bar{X}_j) + u_{0j} + u_{1j} (X_{ij} - \bar{X}_j) + r_{ij}. \]  

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, \( u_{0j} + u_{1j} (X_{ij} - \bar{X}_j) + r_{ij} \). 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 \( u_{0j} + u_{1j} (X_{ij} - \bar{X}_j) \) depend on \( u_{0j} \) and \( u_{1j} \), which vary across schools, and on the value of \( (X_{ij} - \bar{X}_j) \), which varies across students. Though standard regression analysis is inappropriate, such models can be estimated by iterative maximum likelihood procedures described in the

Figure 2.4. Plot of School Means (vertical axis) and SES Slopes (horizontal axis) for 200 Hypothetical Schools. Some students are labeled with a + to indicate that they have a high achievement score.
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 \( u_{ij} \) and \( u_{ij} \) are level-2 random effects. Moreover, \( \text{Var}(r_{ij}) \) is the level-1 variance, and \( \text{Var}(u_{ij}), \text{Var}(u_{ij}), \text{Var}(u_{ij}), \text{Cov}(u_{ij}, u_{ij}), \text{Cov}(u_{ij}, u_{ij}) \) are the level-2 variance-covariance components. The \( \beta \) parameters in the level-1 model are level-1 coefficients and the \( y \)s 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, \( \beta_{ij} \) in the level-1 model is set to zero for all \( j \), yielding

\[
y_{ij} = \beta_{ij} + r_{ij}.
\]  

We assume that each level-1 error, \( r_{ij} \), is normally distributed with a mean of zero and a constant level-1 variance, \( \sigma^2 \). Notice that this model predicts the outcome within each level-1 unit with just one level-2 parameter, the intercept, \( \beta_{00} \). In this case, \( \beta_{00} \) is just the mean outcome for the \( j \)th unit. That is, \( \beta_{00} = \mu_Y \).

The level-2 model for the one-way ANOVA with random effects is Equation 2.4a with \( \gamma_{00} \) set to zero:

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

next chapter. We note that if \( u_{0j} \) and \( u_{ij} \) 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 \( \beta_{0j} \) and \( \beta_{ij} \), 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.
Hierarchical Linear Models

where $\gamma_{00}$ represents the grand-mean outcome in the population, and $\mu_{0j}$ is the random effect associated with unit $j$ and is assumed to have a mean of zero and variance $\tau_{00}$.

Substituting Equation 2.7 into Equation 2.6 yields the combined model

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

[2.8]

which is, indeed, the one-way ANOVA model with grand mean $\gamma_{00}$; 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

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

[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, $\gamma_{00}$. More important, it provides information about the outcome variability at each of the two levels. The $\sigma^2$ parameter represents the within-group variability, and $\tau_{00}$ 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

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

[2.10]

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

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,

$$\beta_{0j} = \gamma_{00} + \gamma_{0}W_{j} + u_{0j},$$

[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:

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

[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

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

Similarly, the variance in $u_{0j}$, $\tau_{00}$, is now the residual or conditional variance in $\beta_{0j}$ 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 $\gamma_{0j}$ and $\gamma_{1j}$ and the random effects $u_{ij}$ (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,

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{..}) + r_{ij}.$$  

[2.13]

The level-2 model becomes

$$\beta_{0j} = \gamma_{00} + u_{0j},$$

[2.14a]

$$\beta_{1j} = \gamma_{10}.$$  

[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

$$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_{..}) + u_{0j} + r_{ij}.$$  

[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, $\gamma_{10}$ is the pooled within-group regression coefficient of $Y_{ij}$ on $X_{ij}$. Each $\beta_{0j}$ is now the mean outcome for each level-2 unit adjusted for differences among these units in $X_{ij}$. Specifically, $\beta_{0j} = \mu_{Y_{ij}} - \gamma_{10}(\bar{X}_{..} - \bar{X}_{..})$, where $\mu_{Y_{ij}}$ is the mean outcome in school $j$. We also note that the $\text{Var}(r_{ij}) = \sigma^2$ is now a residual variance after adjusting for the level-1 covariate, $X_{ij}$.
Hierarchical Linear Models

An extension of the random-effects ANCOVA allows for the introduction of level-2 covariates. For example, if the coefficient \( \gamma_0 \) is nonnull, the combined model becomes

\[
Y_{ij} = \gamma_0 + \gamma_{0j} W_j + \gamma_{10}(X_{ij} - \bar{X}_j) + u_{0j} + r_{ij}. \tag{2.16}
\]

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, \( \gamma_{10} \), 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, \( \beta_{0j} \), 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, \( \beta_{1j} \) 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 \( \gamma_0 \) and \( \gamma_1 \) are constrained to be null. Hence, the level-2 model becomes

\[
\beta_{0j} = \gamma_0 + u_{0j}. \tag{2.17a}
\]
\[
\beta_{1j} = \gamma_{10} + u_{1j}. \tag{2.17b}
\]

The Logic of Hierarchical Linear Models

where

\( \gamma_0 \) is the average intercept across the level-2 units;
\( \gamma_{10} \) 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:

\[
\text{Var} \begin{bmatrix}
\mu_{0j} \\
\mu_{1j}
\end{bmatrix} = \begin{bmatrix}
\tau_{00} & \tau_{01} \\
\tau_{10} & \tau_{11}
\end{bmatrix} = \mathbf{T},
\]

where

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

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 \( \beta_{0j} \) and \( \beta_{1j} \) in Equations 2.17a and 2.17b into Equation 2.2 yields a combined model:

\[
Y_{ij} = \gamma_0 + \gamma_{10}(X_{ij} - \bar{X}_j) + u_{0j} + u_{1j}(X_{ij} - \bar{X}_j) + r_{ij}. \tag{2.19}
\]

This model implies that the outcome \( Y_{ij} \) is a function of the average regression equation, \( \gamma_0 + \gamma_{10}(X_{ij} - \bar{X}_j) \) plus a random error having three components: \( u_{0j} \), the random effect of unit \( j \) on the mean; \( u_{1j}(X_{ij} - \bar{X}_j) \), where \( u_{1j} \) is the random effect of unit \( j \) on the slope \( \beta_{1j} \); 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?”
Hierarchical Linear Models

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 \( X \)'s and of multiple \( W \)'s (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, \( \beta_{1j} \). For example, it might be found that the level-2 predictor \( W_j \) in Equation 2.4b does indeed predict the level-1 slope \( \beta_{1j} \). In fact, the analyst might find that after controlling for \( W_j \), the residual variance of \( \beta_{1j} \) (i.e., the variance of the residuals, \( u_{1ij} \) 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 \( u_{1ij} \) to be zero. This eliminates \( \tau_{11} \), the residual variance of the slope, and \( \tau_{00} \), the residual covariance between the slope and the intercept, as parameters to be estimated.

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

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

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

\[
Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} (X_{ij} - \bar{X}_{.j}) + \gamma_{11} W_j (X_{ij} - \bar{X}_{.j}) + u_{0ij} + r_{ij}. \tag{2.21}
\]

In this model, the slopes do vary from group to group, but their variation is nonrandom. Specifically, as Equation 2.20 shows, the slopes \( \beta_{1j} \) 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 \( \beta_{0j} \) 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, \( \sigma^2 \), and the level-2 variance, \( \tau_{00} \). 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 \( \gamma_{11} W_j (X_{ij} - \bar{X}_{.j}) \). 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

\[
Y_{ij} = \beta_{0j} + \beta_{1j} (X_{1ij} - \bar{X}_{.1j}) + \beta_{2j} (X_{2ij} - \bar{X}_{.2j}) + r_{ij}. \tag{2.22}
\]
Hierarchical Linear Models

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

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

where $\gamma_{20}$ is the common effect of $X_{2j}$ in every level-2 unit. We say that the effect of $\beta_{2j}$ is fixed across level-2 units.

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

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

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

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

According to this model, part of the variation of the slope $\beta_{2j}$ 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 $\beta_{2j}$—that is, $\text{Var}(u_{2j}) = \tau_{22}$—is negligible. Then a model constraining that residual variation to be null would be sensible:

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

In this third case, $\beta_{2j}$ 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 $W_j$s is straightforward. Further, the level-2 model does not need to be identical for each equation. One set of $W_j$s may apply for the intercept, a different set be used for $\beta_{1j}$, another set for $\beta_{2j}$, 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 Logic of Hierarchical Linear Models

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 $\sigma^2_{1j}$, 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 $X$s. We know, for example, that in the simple model

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

...
the intercept, \( \beta_{0j} \), 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 \( \beta_{0j} \), 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 \( \beta_{0j} \) 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., \( \gamma_{00} \) and \( \gamma_{10} \) in Equations 2.4a and 2.4b) depend on the location of the \( W_{ij} \) variables. The numerical stability of estimation is not affected by the location for the \( W_{ij} \)s, but a suitable choice will ease interpretation of results. We describe below some common choices for the location of the \( X \)s and \( W \)s.

**Location of the Xs**

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, \( \beta_{0j} \), will be the expected outcome for a student in school \( j \) who had an SAT of zero. The \( \beta_{0j} \) 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 \( \beta_{0ij} \) is the expected outcome for such a subject. That is, \( \beta_{0ij} = E(Y_{ij} | X_{ij} = 0) \). We wish to emphasize that it is always important to consider the meaning of \( X_{ij} = 0 \) because it determines the interpretation of \( \beta_{0ij} \).

**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

\[
(X_{ij} - \bar{X}_{..}).
\]  

[2.27]

Now, the intercept, \( \beta_{0j} \), is the expected outcome for a subject whose value on \( X_{ij} \) is equal to the grand mean, \( \bar{X}_{..} \). 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 \( j \),

\[
\beta_{0j} = \mu_{1j} - \beta_{1j}(X_{ij} - \bar{X}_{..}).
\]

Similarly, the \( \text{Var}(\beta_{0j}) = \tau_{01} \) 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:

\[
(X_{ij} - \bar{X}_{.,.}).
\]  

[2.28]

In this case, the intercept \( \beta_{0j} \) becomes the unadjusted mean for group \( j \). That is,

\[
\beta_{0j} = \mu_{1j}.
\]  

[2.29]

and \( \text{Var}(\beta_{0j}) \) is now just the variance among the level-2 unit means, \( \mu_{1j} \).

**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 \( \beta_{0j} \) 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

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

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 \( \beta_{0j} \) 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, \( \bar{X}_.. \). This centered predictor can take on two values. If the subject is female, \( X_{ij} - \bar{X}_.. \) will equal the proportion of male students in the sample. If the subject is male, \( X_{ij} - \bar{X}_.. \) will 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, \( \beta_{0j} \), 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, \( X_{ij} - \bar{X}_{.j} \) will take on the value equal to the proportion of male students in school \( j \); for males, \( X_{ij} - \bar{X}_{.j} \) 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 \( \beta_{0j} \) when group-mean centering is employed. The intercept still represents the average outcome for unit \( j, \mu_{0j} \).

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 \( \beta_{0j}, \text{Var}(\beta_{0j}) \), and by implication, all of the covariances involving \( \beta_{0j} \).

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, \( \mathbf{T} \), and random level-1 coefficients, \( \beta_{1j} \). 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 Ws

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 \( \gamma \) 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, \( W_{1j} - \bar{W}_.. \).

Summary of Terms and Notation

Introduced in This Chapter

A Simple Two-Level Model

Hierarchical form:

- Level 1 (e.g., students): \( Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij} \),
- Level 2 (e.g., schools): \( \beta_{0j} = \gamma_{00} + \gamma_{01}W_{j} + u_{0j} \), \( \beta_{1j} = \gamma_{10} + \gamma_{11}W_{j} + u_{1j} \).

Model in combined form:

\[ Y_{ij} = \gamma_{00} + \gamma_{01}X_{ij} + \gamma_{01}W_{j} + \gamma_{11}X_{ij}W_{j} + u_{0j} + u_{1j}X_{ij} + r_{ij}, \]

where we assume:

\[ \mathbb{E}(r_{ij}) = 0, \quad \text{Var}(r_{ij}) = \sigma^2, \]
\[ \mathbb{E} \left[ \begin{array}{c} u_{0j} \\ u_{1j} \end{array} \right] = \mathbf{0}, \quad \text{Var} \left[ \begin{array}{c} u_{0j} \\ u_{1j} \end{array} \right] = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} = \mathbf{T}, \]
\[ \text{Cov}(u_{0j}, r_{ij}) = \text{Cov}(u_{1j}, r_{ij}) = 0. \]
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 \( j \).

\( \beta_{0j}, \beta_{1j} \) are level-1 coefficients. These can be of three forms:

- fixed level-1 coefficients (e.g., \( \beta_{1j} \) in the one-way random-effects ANCOVA model, Equation 2.14b)
- nonrandomly varying level-1 coefficients (e.g., \( \beta_{1j} \) in the nonrandomly-varying-slopes model, Equation 2.20)
- random level-1 coefficients (e.g., \( \beta_{0j} \) and \( \beta_{1j} \) 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])

\( \gamma_0, \ldots, \gamma_{j} \) 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).
\( W_{ij} \) 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.
\( \sigma^2 \) is the level-1 variance.
\( \tau_{00}, \tau_0, \tau_{11} \) 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 \( \beta_{0j}, \beta_{1j} \) are the values of the level-2 variances and covariances based on the random-coefficient regression model.

**Conditional or residual variance-covariance** of \( \beta_{0j}, \beta_{1j} \) are the values of the level-2 variances and covariances after level-2 predictors have been added for \( \beta_{0j} \) and \( \beta_{1j} \) (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, \( \beta_{0j} \).

**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**

- \( X_{ij} \) in the natural metric
- \( (X_{ij} - \bar{X}_{..}) \) called grand-mean centering
- \( (X_{ij} - \bar{X}_{.j}) \) called group-mean centering
- \( X_{ij} \) centered at some theoretically chosen location for \( X \)

**Implications for \( \beta_{0j} \)**

- \( \beta_{0j} = E(Y_{ij}|X_{ij} = 0) \)
- \( \beta_{0j} = \mu_{Y_{j}} = \beta_{1j}(X_{ij} - \bar{X}_{..}) \) (i.e., adjusted level-2 means)
- \( \beta_{0j} = \mu_{Y_{j}} \) (i.e., level-2 means)
- \( \beta_{0j} = E(Y_{ij}|X_{ij} = \text{chosen centering location for } X) \)
An Illustration

Introduction

The purpose of this chapter is to illustrate the use of the techniques of estimation and hypothesis testing presented in Chapter 3. We present a series of analyses based on the models introduced in Chapter 2. For each model, we illustrate the estimation of the fixed effects and variance-covariance components, and demonstrate the use of appropriate hypothesis testing procedures for these various parameters. We also introduce through these examples some useful auxiliary description statistics that can be computed based on the maximum likelihood variance-covariance component estimates. Applications of estimation and hypothesis testing procedures for random level-1 coefficients are demonstrated in the last section of this chapter.

The examples presented below use data from a nationally representative sample of U.S. public and Catholic high schools. These data are a subsample from the 1982 High School and Beyond (HS&B) Survey, and include information on 7,185 students nested within 160 schools: 90 public and 70 Catholic. Sample sizes averaged about 45 students per school.

Attention is restricted to two student-level variables: (a) the outcome, $Y_{ij}$, a standardized measure of math achievement; and (b) one predictor, $(\text{SES})_{ij}$, student socioeconomic status, which is a composite of parental education, parental occupation, and parental income. School-level variables include $(\text{SECTOR})_{i}$, an indicator variable taking on a value of one for Catholic schools and zero for public schools, and $(\text{MEAN SES})_{i}$, the average of the student SES values within each school. In the language introduced in Chapter 2, the level-1 units are students and the level-2 units are schools. $(\text{SES})_{ij}$ is a level-1 predictor; $(\text{SECTOR})_{i}$ and $(\text{MEAN SES})$ are level-2 predictors. Means and standard deviations of these variables are supplied in Table 4.1.

Questions motivating these analyses include the following:

1. How much do U.S. high schools vary in their mean mathematics achievement?
2. Do schools with high MEAN SES also have high math achievement?
3. Is the strength of association between student SES and math achievement similar across schools? Or is SES a more important predictor of achievement in some schools than in others?
4. How do public and Catholic schools compare in terms of mean math achievement and in terms of the strength of the SES-math achievement relationship, after we control for MEAN SES?

The One-Way ANOVA

The one-way ANOVA with random effects, described in Chapter 2, provides useful preliminary information about how much variation in the outcome lies within and between schools and about the reliability of each school's sample mean as an estimate of its true population mean.

The Model

The level-1 or student-level model is

$$Y_{ij} = \beta_{0j} + r_{ij}, \quad [4.1]$$
where we assume \( r_{ij} \sim \text{iid } N(0, \tau^2) \) for \( i = 1, \ldots, n_j \) students in school \( j \), and \( j = 1, \ldots, 160 \) schools. We refer to \( \tau^2 \) as the student-level variance. Notice that this model characterizes achievement in each school with just an intercept, \( \beta_0j \), which in this case is the mean.

At level 2 or the school level, each school’s mean math achievement, \( \beta_0j \), is represented as a function of the grand mean, \( \gamma_00 \), plus a random error, \( u_{0j} \):

\[
\beta_{0j} = \gamma_{00} + u_{0j}, \tag{4.2}
\]

where we assume \( u_{0j} \sim \text{iid } N(0, \tau_{00}) \). We refer to \( \tau_{00} \) as the student-level variance.

This yields a combined model, also often referred to as a mixed model, with fixed effect \( \gamma_{00} \) and random effects \( u_{0j} \) and \( r_{ij} \):

\[
Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}. \tag{4.3}
\]

**Results**

**Fixed Effects.** From Table 4.2, the weighted least squares estimate for the grand-mean math achievement (using the estimator from Equation 3.9) is

\[
\hat{\gamma}_{00} = 12.64.
\]

This has a standard error of 0.24 and yields a 95% confidence interval (see Equation 3.12) of

\[
12.64 \pm 1.96(0.24) = (12.17, 13.11).
\]

**Variance Components.** Table 4.2 also lists restricted maximum likelihood estimates of the variance components. At the student level,

\[
\hat{\text{Var}}(r_{ij}) = \hat{\sigma}^2 = 39.15.
\]

**TABLE 4.2 Results from the One-Way ANOVA Model**

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average school mean, ( \gamma_{00} )</td>
<td>12.64</td>
<td>0.24</td>
</tr>
</tbody>
</table>

**Random Effect**

<table>
<thead>
<tr>
<th>Variance Component</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>( p ) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>School mean, ( u_{0j} ), ( \tau_{0j} )</td>
<td>8.61</td>
<td>159</td>
<td>1.6602</td>
</tr>
<tr>
<td>Level-1 effect, ( r_{ij} ), ( \sigma^2 )</td>
<td>39.15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

At the school level, \( \tau_{00} \) is the variance of the true school means, \( \beta_0j \), around the grand mean, \( \gamma_{00} \). The estimated variability in these school means is

\[
\hat{\text{Var}}(\beta_{0j}) = \hat{\text{Var}}(u_{0j}) = \hat{\tau}_{00} = 8.61, \tag{4.4}
\]

To gauge the magnitude of the variation among schools in their mean achievement levels, it is useful to calculate the plausible values range for these means. Under the normality assumption of Equation 4.2, we would expect 95% of the school means to fall within the range:

\[
\gamma_{00} \pm 1.96(\hat{\tau}_{00})^{1/2}, \tag{4.4}
\]

which yields

\[
12.64 \pm 1.96(8.61)^{1/2} = (6.89, 18.39).
\]

This indicates a substantial range in average achievement levels among schools in this sample of data.

We may wish to test formally whether the estimated value of \( \tau_{00} \) is significantly greater than zero. If not, it may be sensible to assume that all schools have the same mean. Formally, this hypothesis is

\[
H_0: \tau_{00} = 0,
\]

which may be tested using Equation 3.103. This test statistic reduces in a one-way random ANOVA model to

\[
H = \sum n_j (\bar{Y}_{.j} - \hat{\gamma}_{00})^2 / \hat{\sigma}^2, \tag{4.5}
\]

which has a large-sample \( \chi^2 \) distribution with \( J - 1 \) degrees of freedom under the null hypothesis. In our case, the test statistic takes on a value of 1,660.2 with 159 degrees of freedom (\( J = 160 \) schools). The null hypothesis is highly implausible (\( p < .001 \)), indicating significant variation does exist among schools in their achievement.

**Auxiliary Statistics.** The intraclass correlation, which represents in this case the proportion of variance in \( Y \) between schools, is estimated by substituting the estimated variance components for their respective parameters in Equation 2.10:

\[
\hat{\rho} = \hat{\tau}_{00} / (\hat{\tau}_{00} + \hat{\sigma}^2) = 8.61 / (8.61 + 39.15) = 0.18, \tag{4.6}
\]

indicating that about 18% of the variance in math achievement is between schools.
Similarly, an estimator of the reliability of the sample mean in any school for the true school mean, \( \beta_{0j} \), can also be derived by substituting the estimated variance components into Equation 3.42. That is,

\[
\hat{\lambda}_j = \text{reliability}(\bar{Y}_{.j}) = \frac{\tau_{00}}{\[\frac{\tau_{00}}{n_j} + (\sigma^2/n_j)\]},
\]

In general, the reliability of the sample mean \( \bar{Y}_{.j} \) will vary from school to school because the sample size, \( n_j \), varies. However, an overall measure of the reliability can be obtained by averaging the individual school estimates:

\[
\hat{\lambda} = \frac{\sum \hat{\lambda}_j}{J}.
\]

For the HS&B data, \( \hat{\lambda} = .90 \), indicating that the sample means tend to be quite reliable as indicators of the true school means.

In summary, this one-way ANOVA produces useful preliminary information in our study of math achievement in U.S. high schools. It provides an estimate of the grand mean; a partitioning of the total variation in math achievement into variation between and within schools; a range of plausible values for the school means and a test of the hypothesis that the variability is null; information on the degree of dependence of the observations within each school (the intraclass correlation); and a measure of the reliability of each school’s sample average math achievement as an estimate of its true mean.

Regression with Means-as-Outcomes

The Model

The student-level model of Equation 4.1 remains unchanged: Student math achievement scores are viewed as varying around their school means. The school-level model of Equation 4.2 is now elaborated, however, so that each school’s mean is now predicted by the MEAN SES of the school:

\[
\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{MEAN SES})_j + u_{0j},
\]

where \( \gamma_{00} \) is the intercept, \( \gamma_{01} \) is the effect of MEAN SES on \( \beta_{0j} \), and we assume \( u_{0j} \sim \text{independently } N(0, \tau_{00}) \).

Notice that the symbols \( u_{0j} \) and \( \tau_{00} \) have different meanings in Equation 4.2. Whereas the random variable \( u_{0j} \) had been the deviation of school \( j \)'s mean from the grand mean, it now represents the residual \( \beta_{0j} - \gamma_{00} - \gamma_{01}(\text{MEAN SES})_j \). Correspondingly, the variance \( \tau_{00} \) is now a

 residual or conditional variance, that is, \( \text{Var}(\beta_{0j}|\text{MEAN SES}) \), the school-level variance in \( \beta_{0j} \) after controlling for school MEAN SES.

Substituting Equation 4.9 into Equation 4.1 yields the combined model (or “mixed model”)

\[
Y_{ij} = \gamma_{00} + \gamma_{01}(\text{MEAN SES})_j + u_{0j} + e_{ij}
\]

having fixed effects \( \gamma_{00}, \gamma_{01} \), and random effects \( u_{0j}, e_{ij} \).

Results

Table 4.3 provides estimates and hypothesis tests for the fixed effects and the variances of the random effects.

**Fixed Effects.** We see a highly significant association between school MEAN SES and mean achievement (\( \hat{\gamma}_{01} = 5.68, t = 16.22 \)). The \( t \) ratio employed for hypothesis testing of an individual fixed effect is simply the ratio of the estimated coefficient to its standard error (see Equation 3.83):

\[
t = \frac{\hat{\gamma}_{01}}{\sqrt{\text{Var}(\hat{\gamma}_{01})}} = 5.86/0.36 = 16.22.
\]

It is also possible to test the hypothesis that the grand mean is null, that is, \( H_0: \gamma_{00} = 0 \), but that hypothesis is of no interest in this case.

**Variance Component.** The residual variance between schools, \( \hat{\tau}_{00} = 2.64 \), is substantially smaller than the original, \( \hat{\tau}_{00} = 8.61 \), estimated in the context of the random ANOVA model (see Table 4.2). A range of plausible values for school means, given that all schools have a MEAN SES of zero, is

\[
\hat{\gamma}_{00} \pm 1.96(\hat{\tau}_{00})^{1/2} = 12.65 \pm 1.96(2.64)^{1/2} = (9.47, 15.83).
\]

<table>
<thead>
<tr>
<th>TABLE 4.3 Results from the Means-as-Outcomes Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed Effect</strong></td>
</tr>
<tr>
<td>Model for school means</td>
</tr>
<tr>
<td>INTERCEPT, ( \gamma_{00} )</td>
</tr>
<tr>
<td>MEAN SES, ( \gamma_{01} )</td>
</tr>
<tr>
<td>Random Effect</td>
</tr>
<tr>
<td>School mean, ( u_{0j} )</td>
</tr>
<tr>
<td>Level-1 effect, ( r_{ij} )</td>
</tr>
</tbody>
</table>
Though this is a fairly wide range of plausible values, it is considerably smaller than the range of plausible values when MEAN SES is not held constant (Equation 4.4), which was (6.89, 18.39).

Do school achievement means vary significantly once MEAN SES is controlled? Here the null hypothesis that $\tau_{00} = 0$, where $\tau_{00}$ is now a residual variance, is tested by means of the statistic

$$\sum \eta_{ij}(\bar{Y}_{ij} - \hat{\gamma}_{00} - \hat{\gamma}_{01}(\text{MEAN SES})_{i})^2 / \hat{\sigma}^2,$$

which, under the null hypothesis, has a $\chi^2$ distribution with $J - 2 = 158$ degrees of freedom. In our case, the statistic has a value of 633.52, $p < .001$, indicating that the null hypothesis is easily rejected; after controlling for MEAN SES, significant variation among school mean math achievement remains to be explained.

**Auxiliary Statistics.** By comparing the $\tau_{00}$ estimates across the two models, we can develop an index of the proportion reduction in variance or "variance explained" at level 2. In this application,

$$\text{Proportion of variance explained in } \beta_{0ij} = \frac{\hat{\tau}_{00}(\text{random ANOVA}) - \hat{\tau}_{00}(\text{MEAN SES})}{\hat{\tau}_{00}(\text{random ANOVA})},$$

where $\hat{\tau}_{00}(\text{random ANOVA}) = \text{Var}(\beta_{0ij})$ and $\hat{\tau}_{00}(\text{MEAN SES}) = \text{Var}(\beta_{0ij})$ MEAN SES refer to the estimates of $\tau_{00}$ under the alternative level-2 models specified by Equations 4.2 and 4.9, respectively. Note the $\hat{\tau}_{00}(\text{random ANOVA})$ provides the base in this application, because it represents the total parameter variance in the school means that is potentially explainable by alternative level-2 models for $\beta_{0ij}$. The estimated proportion of variance between schools explained by the model with MEAN SES is

$$(8.61 - 2.64)/8.61 = 0.69.$$ 

That is, 69% of the true between-school variance in math achievement is accounted for by MEAN SES.

After removing the effect of school MEAN SES, the correlation between pairs of scores in the same school, which had been .18, is now reduced:

$$\hat{\rho} = \hat{\tau}_{00} / (\hat{\tau}_{00} + \hat{\sigma}^2)$$

$$= 2.64/(2.64 + 39.16) = .06.$$ 

*An Illustration*

The estimated $\rho$ is now a conditional intraclass correlation and measures the degree of dependence among observations within schools that are of the same MEAN SES.

Similarly, we can calculate the reliability of the least squares residuals, $\hat{\eta}_{0ij}$,

$$\hat{\eta}_{0ij} = \bar{Y}_{ij} - \hat{\gamma}_{00} - \hat{\gamma}_{01}(\text{MEAN SES})_{i}.$$

This reliability is a conditional reliability, that is, the reliability with which one can discriminate among schools that are identical on MEAN SES. Substituting our new estimates of $\hat{\tau}_{00}$ and $\hat{\sigma}^2$ into Equations 4.7 and 4.8 yields an average reliability of .74. As one might expect, the reliability of the residuals is less than the reliability of the sample means.

In summary, we have learned from the means-as-outcomes model that MEAN SES is significantly positively related to mean achievement. Nonetheless, even after we hold constant, or control for, MEAN SES, schools still vary significantly in their average achievement levels.

**The Random-Coefficient Model**

We now consider an analysis of the SES-math achievement relationship within the 160 schools. We conceive of each school as having "its own" regression equation with an intercept and a slope, and we shall ask the following:

1. What is the average of the 160 regression equations (i.e., what are the average intercept and slope)?
2. How much do the regression equations vary from school to school? Specifically, how much do the intercepts vary and how much do the slopes vary?
3. What is the correlation between the intercepts and the slopes? (Do schools with large intercepts [e.g., high mean achievement] also have large slopes [strong relationships between SES and achievement]?)

**The Model**

To answer these questions, we use the random-coefficient regression model introduced in Chapter 2. Specifically, we formulate at level 1 the student-level model

$$Y_{ij} = \beta_{0ij} + \beta_{1ij}(X_{ij} - \bar{X}_{.}) + r_{ij}.$$ 

Each school's distribution of math achievement is characterized by two parameters: the intercept, $\beta_{0ij}$, and the slope, $\beta_{1ij}$. Because the student-level
predictor is centered around its school mean, the intercept, $\beta_{0j}$, is the school-mean outcome (see Equation 2.29). Again, we assume $r_{ij} \sim \text{N}(0, \sigma^2)$, where now $\sigma^2$ is the residual variance at level 1 after controlling for student SES.

These parameters, $\beta_{0j}$ and $\beta_{1j}$, vary across schools in the level-2 model as a function of a grand mean and a random error:

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad [4.15a]$$

$$\beta_{1j} = \gamma_{10} + u_{1j}, \quad [4.15b]$$

where

$\gamma_{00}$ is the average of the school means on math achievement across the population of schools;

$\gamma_{10}$ is the average SES-math regression slope across those schools;

$u_{0j}$ is the unique increment to the intercept associated with school $j$; and

$u_{1j}$ is the unique increment to the slope associated with school $j$.

Substituting Equation 4.15 into Equation 4.14 yields a mixed model:

$$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \overline{X}_j) + u_{0j} + u_{1j}(X_{ij} - \overline{X}_j) + r_{ij}. \quad [4.16]$$

We assume that $u_{0j}$ and $u_{1j}$ are multivariate normally distributed, both with expected values of $0$. We label the variances in these school effects as

$$\text{Var}(u_{0j}) = \tau_{00},$$

$$\text{Var}(u_{1j}) = \tau_{11},$$

and the covariance between them as

$$\text{Cov}(u_{0j}, u_{1j}) = \tau_{01}.$$

Collecting these terms into a variance-covariance matrix,

$$\text{Var} \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} = T. \quad [4.17]$$

Because the level-2 model is unconditional for both $\beta_{0j}$ and $\beta_{1j}$ (i.e., no predictors are included in Equations 4.15a and 4.15b),

$$\text{Var}(u_{0j}) = \text{Var}(\beta_{0j} - \gamma_{00}) = \text{Var}(\beta_{0j}),$$

$$\text{Var}(u_{1j}) = \text{Var}(\beta_{1j} - \gamma_{10}) = \text{Var}(\beta_{1j}). \quad [4.18]$$

Thus, the random-coefficient regression model provides estimates for the unconditional parameter variability in the random intercepts and slopes.

### TABLE 4.4 Results from the Random-Coefficient Model

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>se</th>
<th>t Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall mean achievement, $\gamma_{00}$</td>
<td>12.64</td>
<td>0.24</td>
<td>—</td>
</tr>
<tr>
<td>Mean SES-achievement slope, $\gamma_{10}$</td>
<td>2.19</td>
<td>0.13</td>
<td>17.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Variance</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>School mean, $u_{0j}$</td>
<td>8.68</td>
<td>159</td>
<td>1,770.9</td>
<td>0.000</td>
</tr>
<tr>
<td>SES-achievement slope, $u_{1j}$</td>
<td>0.68</td>
<td>159</td>
<td>213.4</td>
<td>0.003</td>
</tr>
<tr>
<td>Level-1 effect, $r_{ij}$</td>
<td>36.70</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Results

**Fixed Effects.** Table 4.4 provides the estimates for the average regression equation within schools. Using the generalized least squares estimator of Equation 3.31 (or, equivalently, Equation 3.37), we find the average school mean

$$\hat{\gamma}_{00} = 12.64$$

and the average SES-achievement slope

$$\hat{\gamma}_{10} = 2.19.$$  

The corresponding standard errors, based on Equation 3.32 (or Equation 3.38), are 0.24 and 0.13, respectively. We can use this information to formally test the null hypothesis that, on average, student SES is not related to math achievement within schools, that is,

$$H_0: \gamma_{10} = 0.$$

Based on the test statistic of Equation 3.83:

$$t = \frac{\hat{\gamma}_{10}}{\sqrt{\hat{\Sigma}_{00}^{1/2}}} = 2.19/0.13 = 17.16,$$

we find that, on average, student SES is significantly related ($p < .001$) to math achievement within schools.

**Variance-Covariance Components.** Using the general procedures for maximum likelihood estimation discussed in Chapter 3, we estimate the variance and covariance components of Equation 4.17:

$$\text{Var} \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \hat{\tau}_{00} & \hat{\tau}_{01} \\ \hat{\tau}_{10} & \hat{\tau}_{11} \end{bmatrix} = \begin{bmatrix} 8.68 & 0.04 \\ 0.04 & 0.68 \end{bmatrix} = T.$$
Table 4.4 also provides the test statistics (see Equation 3.101) for the hypotheses that each of the variance components along the diagonal of \( T \) are null, that is, 

\[
H_0: \tau_{qq} = 0 \quad \text{for } q = 0, 1.
\]

Specifically, the estimated variance among the means is \( \hat{\tau}_{00} = 8.68 \), with a \( \chi^2 \) statistic of 1,770.5, to be compared to the critical value of \( \chi^2 \) with \( J - 1 = 159 \) degrees of freedom. We infer that highly significant differences exist among the 160 school means, a result quite similar to that encountered in the one-way ANOVA with random effects.

The estimated variance of the slopes is \( \hat{\tau}_{11} = 0.68 \) with a \( \chi^2 \) statistic of 213.4 and 159 degrees of freedom, \( p < .003 \). Again, we reject the null hypothesis, in this case that \( \tau_{11} = 0 \), and infer that the relationship between SES and math achievement within schools does indeed vary significantly across the population of schools.

We can also now calculate a range of plausible values for both the school means and the school-specific SES-achievement slopes. Generalizing from Equation 4.4, the 95% plausible value range is

\[
\hat{\tau}_{qq} \pm 1.96(\hat{\tau}_{qq})^{1/2}
\]  

[4.19]

for the \( q = 0, \ldots, Q \) random coefficients in the level-1 model. (In this example, \( Q = 1 \).) The 95% plausible value range for the school means is

\[
12.64 \pm 1.96(8.68)^{1/2} = (6.87, 18.41),
\]

and for the SES-achievement slopes is

\[
2.19 \pm 1.96(0.68)^{1/2} = (0.57, 3.81).
\]

The results for the school means are similar to those previously reported for the one-way ANOVA model. In terms of the school-specific SES-achievement slopes, here, too, we find considerable variability among schools. This relationship is over seven times stronger in the most socially differentiating schools as compared to the least differentiating schools.

**Auxiliary Statistics.** Associated with \( \beta_{0j} \) and \( \beta_{1i} \) is also a reliability estimate (see Equation 3.59). The results are 

\[
\text{reliability}(\hat{\beta}_0) = 0.91
\]

and

\[
\text{reliability}(\hat{\beta}_1) = 0.23.
\]

These indices provide answers to the question "How reliable, on average, are estimates of each school’s intercept and slope based on computing the OLS regression separately for each school?" These reliabilities depend on two factors: the degree to which the true underlying parameters vary from school to school and the precision with which each school’s regression equation is estimated.

The precision of estimation of the intercept (which in this application is a school mean) depends on the sample size within each school. The precision of estimation of the slope for school \( j \) depends both on the sample size and on the variability of SES within that school. Schools that are homogeneous with respect to SES will exhibit slope estimates with poor precision.

The results indicate that the intercepts are quite reliable (.91) based on an average of 50 students per school. The slope estimates are far less reliable (.23). The primary reason for the lack of reliability of the slopes is that the true slope variance across schools is much smaller than the variance of the true means. Also, the slopes are estimated with less precision than are the means because many schools are relatively homogeneous on SES.

Analogous to Equation 4.12, we can develop an index of the proportion reduction in variance or "variance explained" at level 1 by comparing the \( \sigma^2 \) estimates from these two alternative models. Notice that the estimate of the student-level variance \( \hat{\sigma}^2 \) is now 36.70. By comparison, the estimated variance in the one-way random ANOVA model, which did not include SES as a level-1 predictor, was 39.15. Thus,

\[
\text{Proportion variance explained at level 1} = \frac{\hat{\sigma}^2(\text{random ANOVA}) - \hat{\sigma}^2(\text{SES})}{\hat{\sigma}^2(\text{random ANOVA})} = \frac{39.15 - 36.70}{39.15} = .063,
\]

[4.20]

where \( \hat{\sigma}^2(\text{random ANOVA}) \) and \( \hat{\sigma}^2(\text{SES}) \) refer to estimates of \( \sigma^2 \) based on the level-1 models specified by Equations 4.1 and 4.14, respectively. Note that \( \sigma^2(\text{random ANOVA}) \) provides the appropriate base in this application because it represents the total within-school variance that can be explained by any level-1 model.

We see that adding SES as a predictor of math achievement reduced the within-school variance by 6.3%. Hence, we can conclude that SES accounts for about 6% of the student-level variance in the outcome. When we recall that MEAN SES accounted for better than 60% of the between-school variance in the outcome, it is clear that the association between these two variables is far stronger at the school level than at the student level.
Finally, the model also produces a maximum likelihood estimate of the covariance between the intercept and the slope. When combined with the estimates of the intercept and slope variances, we can estimate the correlation between the intercept and slope using Equation 2.3. In this case, the correlation of slope and intercept is .02, indicating that there is little association between school means and school SES effects.

An Intercepts- and Slopes-as-Outcomes Model

Having estimated the variability of the regression equations across schools, we now seek to build an explanatory model to account for this variability. That is, we seek to understand why some schools have higher means than others and why in some schools the association between SES and achievement is stronger than in others.

The Model

The student-level model remains the same as in Equation 4.14. However, we now expand the student-level model to incorporate two predictors: SECTOR and MEAN SES. The resulting school-level model can be written as

\[
\begin{align*}
\beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{MEAN SES})_j + \gamma_{02}(\text{SECTOR})_j + u_{0j}, \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}(\text{MEAN SES})_j + \gamma_{12}(\text{SECTOR})_j + u_{1j},
\end{align*}
\]

where \(u_{0j}\) and \(u_{1j}\) are again multivariate normally distributed with means of zero and variance-covariance matrix \(T\). The elements of \(T\) are now residual or conditional variance-covariance components. That is, they represent residual dispersion in \(\beta_{0j}\) and \(\beta_{1j}\) after controlling for MEAN SES and SECTOR.

Combining the school-level model (Equation 4.21) and the student-level model (Equation 4.14) yields

\[
Y_{ij} = \gamma_{00} + \gamma_{01}(\text{MEAN SES})_j + \gamma_{02}(\text{SECTOR})_j + \gamma_{10}(X_{ij} - \bar{X}_j) + \gamma_{11}(\text{MEAN SES})_j(X_{ij} - \bar{X}_j) \\
+ \gamma_{12}(\text{SECTOR})_j(X_{ij} - \bar{X}_j) + u_{0j} + u_{1j}(X_{ij} - \bar{X}_j) + r_{ij},
\]

which illustrates that the outcome may be viewed as a function of the overall intercept (\(\gamma_{00}\)), the main effect of MEAN SES (\(\gamma_{01}\)), the main effect of SECTOR (\(\gamma_{02}\)), the main effect of SES (\(\gamma_{10}\)), and two cross-level interactions involving SECTOR with student SES (\(\gamma_{12}\)) and MEAN SES with student SES (\(\gamma_{11}\)), plus a random error

\[
u_{0j} + u_{1j}(X_{ij} - \bar{X}_j) + r_{ij}.
\]

Three kinds of questions motivate the analysis:

1. Do MEAN SES and SECTOR significantly predict the intercept? We estimate \(\gamma_{01}\) to study whether high-SES schools differ from low-SES schools in mean achievement (controlling for SECTOR). Similarly, we estimate \(\gamma_{02}\) to learn whether Catholic schools differ from public schools in terms of the mean achievement once MEAN SES is controlled.

2. Do MEAN SES and SECTOR significantly predict the within-school slopes? We estimate \(\gamma_{11}\) to discover whether high-SES schools differ from low-SES schools in terms of the strength of association between student SES and achievement within them (controlling for SECTOR). We estimate \(\gamma_{12}\) to examine whether Catholic schools differ from public schools in terms of the strength of association between student SES and achievement (controlling for MEAN SES).

3. How much variation in the intercepts and the slopes is explained by using SECTOR and MEAN SES as predictors? To answer these questions, we estimate \(\text{Var}(u_{0j}) = \tau_{00}\) and \(\text{Var}(u_{1j}) = \tau_{11}\) and compare these with the estimates presented above from the random-coefficient regression model.

Results

Fixed Effects. Table 4.5 displays the results. We see, first, that MEAN SES is positively related to school mean math achievement, \(\hat{\gamma}_{01} = 5.33, t = 14.45\). Also, Catholic schools have significantly higher mean achievement than do public schools, controlling for the effect of MEAN SES, \(\hat{\gamma}_{02} = 1.23, t = 4.00\).

With regard to the slopes, there is a tendency for schools of high MEAN SES to have larger slopes than do schools with low MEAN SES, \(\hat{\gamma}_{11} = 1.03, t = 3.42\). Catholic schools have significantly weaker SES slopes, on average, than do public schools, \(\hat{\gamma}_{12} = -1.64, t = -6.76\).

These results are depicted graphically in Figure 4.1. The fitted relationship between SES and math achievement is displayed for the Catholic and public sectors. Within each sector, results are displayed for (1) a high-SES school (one standard deviation above the mean), (2) a medium-SES school, and (3) a low-SES school (one standard deviation below the mean). Perhaps the most notable feature of the figure is that the within-school math-SES slopes are substantially less steep in the Catholic sector than in the public sector. This sector effect holds for schools at each level of MEAN SES. There is also
a tendency for high-SES schools to have steeper slopes than do low-SES schools. This tendency is evident in both sectors. Main effects of MEAN SES and SECTOR are also evident. The MEAN SES effect is manifest by the solid lines that in both plots have positive slopes.

Chapter 3 discusses a procedure for testing multiparameter hypotheses regarding the fixed effects. One may wonder, for example, whether the variable SECTOR is needed in the model. Perhaps no distinction is justified between Catholic and public schools in terms of effectiveness or equity. The null hypothesis may be written as

\[ H_0: \gamma_{12} = 0 \]

\[ H_0: \gamma_{12} = 0. \]

If \( \gamma_{12} = 0 \), Catholic and public schools do not differ in mean achievement after controlling for MEAN SES. Similarly, if \( \gamma_{12} = 0 \), Catholic and public schools do not differ with respect to their average SES-math achievement relationships with schools. If both null hypotheses are true, the variable SECTOR may be dropped from the model. Using Equation 3.91, we obtain a \( \chi^2 \) statistic of 64.38, \( df = 2, p < .001 \), indicating that one or both of the null hypotheses is false.

**Variance-Covariance Components.** Recall that the results of fitting the random-coefficient model provided information about the variation and

---

**TABLE 4.5 Results from the Intercepts- and Slopes-as-Outcomes Model**

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Coefficient</th>
<th>se</th>
<th>t Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for school means</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERCEPT, 12.10</td>
<td></td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>MEAN SES, 5.33</td>
<td></td>
<td>0.37</td>
<td>14.45</td>
</tr>
<tr>
<td>SECTOR, 1.23</td>
<td></td>
<td>0.31</td>
<td>4.00</td>
</tr>
<tr>
<td>Model for SES-achievement slopes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERCEPT, 2.94</td>
<td></td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>MEAN SES, 1.03</td>
<td></td>
<td>0.30</td>
<td>3.42</td>
</tr>
<tr>
<td>SECTOR, -1.64</td>
<td></td>
<td>0.24</td>
<td>-6.76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Variance Component</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>( p ) Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>School mean, ( u_{0j} )</td>
<td>2.38</td>
<td>157</td>
<td>605.30</td>
<td>0.000</td>
</tr>
<tr>
<td>SES-achievement slope, ( u_{ij} )</td>
<td>0.15</td>
<td>157</td>
<td>162.31</td>
<td>0.369</td>
</tr>
<tr>
<td>Level-1 effect, ( r_{ij} )</td>
<td>36.68</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Figure 4.1.** Regressions of Mathematics Achievement as a Function of Student and School SES Within Catholic and Public Sectors

**NOTE:** Schools 1, 2, and 3 are of high, medium, and low SES, respectively.

Covariation of the intercepts and SES slopes across schools. Our interest now focuses on the residual variation and covariation of the intercepts and slopes, that is, the variation left unexplained by SECTOR and MEAN SES. The maximum likelihood point estimates are

\[
\hat{\beta}_{10} = \left( \hat{\beta}_{00}, \hat{\beta}_{01}, \hat{\beta}_{11} \right) = \left( \frac{0.23}{0.19}, \frac{0.19}{0.15} \right).
\]

We note that both \( \hat{\beta}_{00} \) and \( \hat{\beta}_{11} \), the estimated variances of the intercepts and the slopes, are considerably smaller than they had been without control for SECTOR and MEAN SES. Perhaps there remains no significant residual variation in the intercepts and slopes after control for SECTOR and MEAN SES. Regarding the intercepts, the null hypothesis

\[ H_0: \tau_{00} = 0 \]

is rejected, as indicated by the \( \chi^2 \) statistic of 605.30, \( df = J - S_g - 1 = 157, p < .001 \). Thus, significant variation in the intercepts remains
unexplained even after controlling for SECTOR and MEAN SES. Regarding the slopes, the null hypothesis

\[ H_0: \tau_{1i} = 0 \]

is retained, as indicated by the \( \chi^2 \) statistic of 162.31, df = \( J - \hat{s}_k - 1 = 157 \), \( \rho = .369 \). This test suggests that no significant variation in the slopes remains unexplained after controlling for SECTOR and MEAN SES.

Another way to examine the significance of variance and covariance components is to compare two models, one model including the variance components of interest and a second, simpler model that constrains certain components to zero. If the fit of the simpler model to the data is significantly worse than the fit of the more complex model, the simpler model is rejected as inadequately representing the variation in the data. However, if there is no significant difference in the fit of the two models, the simpler model will typically be preferred. Applying this logic to the current data, we compare a model with random intercepts and slopes to a model with only a random intercept.

Our “intercepts- and slopes-as-outcomes” model included four unique variance and covariance parameters: (a) the student-level variance, \( \sigma^2 \); (b) the residual variance of the school means, \( \tau_{00} \); (c) the residual variance of the SES-math slopes, \( \tau_{11} \); and (d) the residual covariance between the means and the slopes, \( \tau_{01} \). In general, the number of variance-covariance parameters estimated in a two-level model is \( m(m + 1)/2 + 1 \), where \( m \) is the number of random effects in the level-2 model. In our case, \( m = 2 \). Chapter 3 (see Equations 3.105 through 3.107) described a likelihood-ratio test that can be used to test the composite null hypothesis

\[ H_0: \begin{pmatrix} \tau_{1i} = 0 \\ \tau_{0i} = 0 \end{pmatrix} \]

One first estimates the full model with four variance-covariance parameters, and then one estimates a reduced model with just two parameters (\( \sigma^2 \) and \( \tau_{00} \))—that is, where \( \tau_{1i} \) and \( \tau_{0i} \) have been constrained to zero. One then compares the deviance associated with the two models and asks whether the reduction in deviance associated with the more complex model is justified.

In our case, the results are as follows:

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of Parameters</th>
<th>Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted, ( D_0 )</td>
<td>2</td>
<td>46,514.0</td>
</tr>
<tr>
<td>Unrestricted, ( D_1 )</td>
<td>4</td>
<td>46,513.1</td>
</tr>
</tbody>
</table>

The reduction in deviance is 0.9, which is not significant when compared against the \( \chi^2 \) distribution with 2 df. Hence, the simpler model seems justified. We infer that explanatory power is not significantly enhanced by specifying the residual SES-achievement slopes as random. The reduced model with \( \beta_{1j} \) specified as nonrandomly varying appears sufficient.

Auxiliary Statistics. Analogous to Equation 4.12, we can develop a proportion reduction in variance or variance-explained statistic for each of the random coefficients (intercepts and slopes) from the level-1 model. The variance estimates from the random-coefficient regression model estimated earlier provide the base for these statistics:

\[
\text{Proportion variance explained in } \beta_{qj} = \frac{\hat{\tau}_{qq}(\text{random regression}) - \hat{\tau}_{qq}(\text{fitted model})}{\hat{\tau}_{qq}(\text{random regression})}. \tag{4.24}
\]

where \( \hat{\tau}_{qq}(\text{random regression}) \) denotes the \( qth \) diagonal element of \( \mathbf{T} \) estimated under the random-regression model (Equations 4.14 and 4.15) and \( \hat{\tau}_{qq}(\text{fitted model}) \) denotes the corresponding element in the \( \mathbf{T} \) matrix estimated under an intercepts- and slopes-as-outcomes model (in this case Equations 4.14 and 4.21).

In this application, we see a substantial reduction in variance of the school means once MEAN SES and SECTOR are controlled. Specifically, whereas the unconditional variance of intercepts had been 8.68, the residual variance is now \( \hat{\tau}_{00} = 2.38 \). This means that 73% of the parameter variation in mean achievement, \( \text{Var}(\beta_0) \), has been explained by MEAN SES and SECTOR [i.e., \((8.68 - 2.38)/8.68 = 0.73\)]. Similarly, the residual variance of the slopes is \( \hat{\tau}_{1j} = 0.15 \), which, when compared to the unconditional variance of 0.68, implies a reduction of 78%. Clearly, most of the slope variability is associated with MEAN SES and SECTOR. Once this is controlled, only a small residual portion of variation remains unexplained. Chapter 5 provides a more detailed discussion, with some caveats, of strategies for monitoring explained variance (see “Use of Proportion Reduction in Variance Statistics”).

Estimating the Level-1 Coefficients for a Particular Unit

In this chapter, we have characterized the distribution of achievement in each school in terms of two school-specific parameters: a school’s mean math achievement and a regression coefficient describing the relationship
between SES and math achievement. We have viewed these level-1 coefficients as "random parameters" varying over the population of schools with some of that random variation a function of measured predictors. As mentioned in Chapter 3, we can obtain both point and interval estimates for each random level-1 coefficient. Formally, these are empirical Bayes estimators, also known as shrinkage estimators. These shrinkage estimators may be subdivided into two categories: unconditional and conditional shrinkage estimators. We illustrate each approach and compare these with the OLS estimates.

Ordinary Least Squares

The most obvious strategy for estimating the regression equation for a particular school is simply to fit a separate model to each school’s data by ordinary least squares (OLS). The model for each school might simply be Equation 4.14. Recall that with SES centered around the school mean, the intercept \( \beta_{0j} \) is the mean outcome for that school, and the regression coefficient \( \beta_{1j} \) represents the expected difference in achievement per unit difference in SES within that school. OLS will produce unbiased estimates of these parameters for any school that has at least two cases. Indeed, if the errors of the model are independently normally distributed, the OLS estimates are the unique, minimum-variance, unbiased estimators of these parameters. Nonetheless, the OLS estimates for any given school may not be very accurate as we illustrate below.

Columns 1 and 2 in Table 4.6 present separate OLS estimates for 12 selected schools out of the 160 cases in the HS&B data set. These estimates were calculated for each school using Equation 3.54. Based on these estimates, we might identify case 4 as an especially good school where there is a high average level of achievement, \( \hat{\beta}_{04} = 16.26 \), that is distributed in an equitable fashion, \( \hat{\beta}_{14} = 0.13 \).

Figure 4.2a shows the OLS estimates for all 160 schools. The intercept estimates (vertical axis) are plotted against the slope estimates (horizontal axis). Quite a few schools yield negative estimates of the SES-achievement relationship. Moreover, the apparent dispersion of the OLS estimates greatly exceeds the maximum likelihood estimate of the variance of the true slopes. Earlier, we estimated the variance of the true slope parameters to be 0.68. Yet the sample variance of the OLS slope estimates depicted in Figure 4.2a is 2.66. If we were to define effective and equitable schools as those with large means and small SES-math achievement slopes, we might identify many schools like case 4.

### TABLE 4.6 Comparison of Estimated Level-1 Coefficients for a Sample of HS&B Cases

<table>
<thead>
<tr>
<th>Case</th>
<th>OLS Estimates</th>
<th>Empirical Bayes Estimates</th>
<th>Empirical Bayes Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\beta}_{0j} )</td>
<td>( \hat{\beta}_{1j} )</td>
<td>( \hat{\beta}_{0j}^* )</td>
</tr>
<tr>
<td>4</td>
<td>16.26</td>
<td>0.13</td>
<td>15.62</td>
</tr>
<tr>
<td>15</td>
<td>15.98</td>
<td>2.15</td>
<td>15.74</td>
</tr>
<tr>
<td>17</td>
<td>18.11</td>
<td>0.09</td>
<td>17.41</td>
</tr>
<tr>
<td>22</td>
<td>11.14</td>
<td>-0.78</td>
<td>11.22</td>
</tr>
<tr>
<td>27</td>
<td>13.40</td>
<td>4.10</td>
<td>13.32</td>
</tr>
<tr>
<td>53</td>
<td>9.52</td>
<td>3.74</td>
<td>9.76</td>
</tr>
<tr>
<td>69</td>
<td>11.47</td>
<td>6.18</td>
<td>11.64</td>
</tr>
<tr>
<td>75</td>
<td>9.06</td>
<td>1.65</td>
<td>9.28</td>
</tr>
<tr>
<td>81</td>
<td>15.42</td>
<td>5.26</td>
<td>15.25</td>
</tr>
<tr>
<td>90</td>
<td>12.14</td>
<td>1.97</td>
<td>12.18</td>
</tr>
<tr>
<td>135</td>
<td>4.55</td>
<td>0.25</td>
<td>6.42</td>
</tr>
<tr>
<td>153</td>
<td>10.28</td>
<td>0.76</td>
<td>10.71</td>
</tr>
</tbody>
</table>

Unconditional Shrinkage

In general, the accuracy of OLS regression estimates for any school depends on the sample size within the school, \( n_j \), and the range represented in the level-1 predictor variable, \( X_{ij} \). If \( n_j \) is small, the mean estimate, \( \hat{\beta}_{0j} \), will be imprecise. If a school has a small sample or a restricted range on SES, the slope estimate, \( \hat{\beta}_{1j} \), will also tend to be imprecise. The empirical Bayes (EB) estimates of each school’s regression line takes into account this imprecision in the OLS estimates.

Columns 3 and 4 in Table 4.6 present the EB estimates for the 12 selected schools from HS&B. These estimates, \( \hat{\beta}_{0j}^* \) and \( \hat{\beta}_{1j}^* \), are based on the unconditional level-2 model (Equations 4.14 and 4.15) and calculated using Equation 3.56 (or, equivalently, Equation 3.66). Notice that the EB estimates for case 4 differ substantially from the OLS estimates. The estimated average achievement level has dropped by 0.64 (from 16.26 to 15.62) and the SES-achievement slope has risen from 0.13 to 2.05. While the estimate of the overall achievement level for the school remains relatively high, the equity effect has disappeared. While case 4 looked superior to case 15 in terms of the OLS estimates, these two schools appear indistinguishable in terms of the empirical Bayes estimates (15.62 versus 15.74 for \( \hat{\beta}_{0j}^* \) and 2.05 versus 2.19 for \( \hat{\beta}_{1j}^* \)). The key factor here is the relatively small sample size of only 20 students in case 4. As a result, the OLS estimates for this unit are
not very precise, and the EB estimates are shrunk toward the overall mean achievement, \( \hat{\gamma}_0 = 12.64 \), and the overall mean SES-achievement slope, \( \hat{\gamma}_1 = 2.19 \). Notice that this occurs for all of the cases with relatively small sample sizes (cases 17, 69, 135, and 153), with sample sizes 29, 25, 14, and 19, respectively.

Figure 4.2b displays the empirical Bayes estimates of the intercepts (vertical axis) and the math-SES slopes (horizontal axis) for all 160 schools. Notice that the empirical Bayes slope estimates are much more concentrated around the sample average than are the OLS estimates in Figure 4.2a. Unlike the collection of OLS slope estimates, none of the empirical Bayes estimates is negative. Also, the sample variance of the empirical Bayes slope estimates is only .14, much smaller than the sample variance of the OLS slope estimates (2.66). In fact, the sample variance of the empirical Bayes slope estimates is smaller than the maximum likelihood estimate of the variance of the true slopes (0.68).

We note that these results generalize:

\[
\text{Var}(\tilde{\beta}_j) > \frac{\text{Var}(\beta_j)}{\text{max variability in OLS estimates}} > \text{observed variability in the empirical Bayes estimates}
\]

The fact that the empirical Bayes estimates have less variance than the estimated true variance is an expected result. In general, the shrinkage is slightly exaggerated; empirical Bayes tends to pull the estimates “too far” toward the sample average.

It is also interesting to contrast the slope shrinkage in Figure 4.2, with the results for the achievement intercepts. Recall that the intercepts are just the mean achievement in each school and are much more reliably estimated than are the slopes (.91 versus .23). Given the greater precision of the intercept estimates, we would expect that the empirical Bayes estimator would rely more heavily on this component, and less shrinkage should occur. This result is displayed in Figure 4.2 (compare the vertical axes of Figures 4.2a and 4.2b). Unlike the slopes, where the shrinkage is substantial, the difference between the OLS and the empirical Bayes estimates for the intercepts is only modest. This same pattern can be observed in the results for the 12-school subsample presented in Table 4.6.

In general, the behavior of the empirical Bayes estimator is simpler in the case of random-intercept models than in models that also have random slopes.
Conditional Shrinkage

A tool for increasing accuracy in estimating \( \beta_{0j} \) and \( \beta_{1j} \) is conditional shrinkage. Rather than pulling each OLS regression line toward the grand-mean regression line of \( \bar{y}_{00} \) and \( \bar{y}_{10} \), the OLS regression lines will now be pulled toward a predicted value based on the school-level model.

Unconditional shrinkage of the random-coefficient regression model (i.e., Equation 4.15) yields

\[
\begin{bmatrix}
\hat{\beta}_{0j} \\
\hat{\beta}_{1j}
\end{bmatrix} = \Lambda_j \begin{bmatrix}
\hat{\beta}_{0j} \\
\hat{\beta}_{1j}
\end{bmatrix} + (1 - \Lambda_j) \begin{bmatrix}
\bar{y}_{0j} \\
\bar{y}_{1j}
\end{bmatrix},
\]

(4.25)

where \( \Lambda_j \) is based on the estimation of \( \sigma^2 \) and \( T \) for the model in Equations 4.14 and 4.15. In contrast, the intercepts- and slopes-as-outcomes model (see Equation 3.56 or, equivalently, Equation 3.66) yields conditional shrinkage toward predicted values of \( \beta_{0j} \) and \( \beta_{1j} \). That is,

\[
\begin{bmatrix}
\hat{\beta}_{0j} \\
\hat{\beta}_{1j}
\end{bmatrix} = \Lambda_j \begin{bmatrix}
\bar{y}_{0j} + \bar{y}_{0i} \text{ (MEAN SES)} \\
\bar{y}_{1j} + \bar{y}_{1i} \text{ (SECTOR)}
\end{bmatrix} + (1 - \Lambda_j) \begin{bmatrix}
\bar{y}_{0j} + \bar{y}_{0i} \text{ (MEAN SES)} \\
\bar{y}_{1j} + \bar{y}_{1i} \text{ (SECTOR)}
\end{bmatrix},
\]

(4.26)

where \( \Lambda_j \) is now based on the estimation of \( \sigma^2 \) and \( T \) for the model in Equations 4.14 and 4.21.

As with unconditional shrinkage, the effects of conditional shrinkage can be quite extreme when the within-group sample size, \( n_j \), is small. (See results in columns 5 and 6 in Table 4.6.) Notice for case 135 that the OLS estimates for \( \beta_{0j} \) and \( \beta_{1j} \) were (4.55, 0.25) respectively; with unconditional shrinkage, they moved to (6.42, 1.93), respectively; and for conditional shrinkage they ended up at (8.55, 2.61). Under conditional shrinkage, the OLS regression line for case 135 is being pulled toward a predicted value for \( \beta_{0j} \) and \( \beta_{1j} \) based on the information from this school about its social class (MEAN SES = .03) and sector (SECTOR = 0 = public). Substituting these values into the estimated equations in Table 4.5 yields predictions of

\[
\hat{E}(\beta_{0j}) = 12.10 + 5.33(0.03) + 1.23(0) = 12.26;
\]

\[
\hat{E}(\beta_{1j}) = 2.94 + 1.03(0.03) - 1.64(0) = 2.97.
\]

Notice that the empirical Bayes estimate for \( \beta_{1j} \), \( \hat{\beta}_{1j} = 2.61 \), entails virtual complete shrinkage from the original OLS value of 0.25 toward the predicted value of 2.97. In addition to the fact that the sample size is small for this unit (\( n_{135} = 14 \)) and that regression slopes are less reliable, the amount of change under conditional shrinkage also depends on the precision of the prediction equations, which, in turn, are a function of the residual variances in \( T \). Since \( \bar{y}_{11} = 0.15 \), the prediction equation is relatively precise in this application and the shrinkage more extensive. We note that the same factors are at work in the conditional shrinkage of \( \hat{\beta}_{0j} \) toward \( \hat{\beta}_{0j} \). The proportional amount of shrinkage \( \hat{\beta}_{0j} \) versus \( \hat{\beta}_{1j} \) is less, however, because school means are considerably more reliable than slopes, for any fixed value of \( n_j \). Since the original sample mean is relatively more precise, empirical Bayes gives relatively more weight to \( \hat{\beta}_{0j} \) in the estimation of \( \hat{\beta}_{0j} \).

Results in Table 4.6 also illustrate the differential effect that conditional shrinkage can have on various units. Compare, for example, case 22 (a low-SES Catholic school) with case 27 (an average-SES public school). The OLS estimates suggest substantial differences between these two schools in both average achievement level (11.14 versus 13.40) and SES-achievement slopes (−0.78 versus 4.10). With unconditional shrinkage, much of the observed differences are “shrunken.” We know, however, from the results in Table 4.6 that school MEAN SES and SECTOR predict both the mean achievement level and the SES-achievement slope for a school. When we take this into account through conditional shrinkage, most of the original differences reappear: 10.89 versus 12.95 for \( \beta_{0j} \) and 0.58 versus 3.02 for \( \beta_{1j} \).

Another interesting set of comparisons are schools 17 and 81. Both are relatively high SES schools, with case 17 being public and case 81 being Catholic. Both schools have relatively high average achievement levels (\( \hat{\beta}_{0j} = 18.11; \hat{\beta}_{0j} = 15.42 \)), but the Catholic school has a very steep SES-achievement slope (\( \hat{\beta}_{1j} = 5.26 \)) as compared to the public school (\( \hat{\beta}_{1j} = 0.09 \)). Both of the schools are multivariate outliers in that their OLS slope estimates appear inconsistent with all the other information we have in the data set. That is, from Table 4.5 we expect high-SES public schools to have steep slopes, not their Catholic counterparts. In this instance, conditional shrinkage literally reorders the two equations. While \( \hat{\beta}_j \) remains higher for the public school (17.25 versus 15.52), the conditional shrinkage estimate for the SES-achievement slopes is now lower in the Catholic school than in the public (2.01 versus 3.67)! The effects of conditional shrinkage can also be discerned through the use of OLS and empirical Bayes residuals (see Equations 3.50, 3.60; and
3.49, 3.61 respectively). The OLS residuals for the intercepts and slopes in Equation 4.21 are

\[ \hat{u}_{0j} = \hat{\beta}_{0j} - \left[ \hat{\gamma}_{00} + \hat{\gamma}_{01}(\text{MEAN SES})_j + \hat{\gamma}_{02}(\text{SECTOR})_j \right]. \] \[ \hat{u}_{1j} = \hat{\beta}_{1j} - \left[ \hat{\gamma}_{10} + \hat{\gamma}_{11}(\text{MEAN SES})_j + \hat{\gamma}_{12}(\text{SECTOR})_j \right]. \]

The corresponding empirical Bayes residuals are

\[ u_{0j} = \beta_{0j} - \left[ \hat{\gamma}_{00} + \hat{\gamma}_{01}(\text{MEAN SES})_j + \hat{\gamma}_{02}(\text{SECTOR})_j \right]. \] \[ u_{1j} = \beta_{1j} - \left[ \hat{\gamma}_{10} + \hat{\gamma}_{11}(\text{MEAN SES})_j + \hat{\gamma}_{12}(\text{SECTOR})_j \right]. \]

Figure 4.3 displays the results. The intercept residuals are plotted on the vertical axis and the slope residuals on the horizontal axis. The OLS slope residuals are highly misleading. They suggest considerable unexplained variability in the SES-achievement relationships. In contrast, the empirical Bayes residuals are tightly clumped around zero with even less dispersion than in Figure 4.2. This result is consistent with the results in Table 4.5, where 78% of the variability in \( \beta_{1j} \) was accounted for by MEAN SES and SECTOR.

In contrast, the empirical Bayes and OLS residuals for the intercept are much more similar. These residuals, however, are less dispersed than for the unconditional model (Figure 4.2), which is consistent with the fact that Equation 4.21a accounts for about 73% of the variance in \( \beta_{0j} \).

Comparison of Interval Estimates

In addition to point estimates for the level-1 coefficients, we can also compute empirical Bayes interval estimates using Equation 3.65. We illustrate here the use of these procedures for two selected schools, cases 22 and 135 from the HS&B data, and compare results to confidence interval estimates from separate OLS regressions on each school’s individual data set. The 95% confidence intervals for \( \beta_{0j} \) and \( \beta_{1j} \) under ordinary least squares, unconditional and conditional shrinkage appear in Table 4.7.

Notice that for case 22, where \( n_j = 67 \), the widths of the confidence intervals for school mean achievement, \( \beta_{0j} \), are quite similar across all three analyses. This results from the large within-school sample size coupled with the overall high reliability for school means. In contrast, case 135, where \( n_j = 14 \), experiences some reduction in the width of the confidence intervals as we move from OLS to unconditional shrinkage to conditional shrinkage. The 95% confidence interval under the conditional model of Equation 4.21 is about a third smaller than obtained from ordinary least squares estimates using only this school’s data. A gain in precision has been effected by bringing to bear on this estimation problem all of the information in the data set.

These improvements in precision are even more substantial when we compare the confidence interval estimates for the SES-achievement slopes. For case 22, the empirical Bayes 95% confidence interval, based on the conditional model, is about 75% smaller than OLS; for case 135, it is about 85% smaller. Notice that in both cases the likelihood of a negative
TABLE 4.7 Comparison of 95% Confidence Interval Estimates for Random Level-1 Coefficients

<table>
<thead>
<tr>
<th>Case</th>
<th>OLS Unconditional Shrinkage</th>
<th>OLS Conditional Shrinkage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unconditional Shrinkage</td>
<td>Conditional Shrinkage</td>
</tr>
<tr>
<td></td>
<td>( \beta_{q} )</td>
<td>( \beta_{q} )</td>
</tr>
<tr>
<td>22</td>
<td>(9.69, 12.59)</td>
<td>(9.81, 12.63)</td>
</tr>
<tr>
<td>135</td>
<td>(1.37, 7.73)</td>
<td>(3.65, 9.21)</td>
</tr>
</tbody>
</table>

SES-achievement slope becomes trivial under the conditional model; in contrast, under OLS, negative results appear quite plausible.

Cautionary Note

The conditional shrinkage estimators will be substantially more accurate than the OLS estimators when the level-2 model is appropriately specified. That is, the underlying assumption of empirical Bayes conditional shrinkage is that, given the predictors in the level-2 model, the regression lines are “conditionally exchangeable.” This means, in the case of Equation 4.21, that once MEAN SES and SECTOR have been taken into account, there is no reason to believe that the deviation of any school’s regression line from its predicted value is larger or smaller than that of any other school. This assumption depends strongly on the validity of the level-2 model. If that model is misspecified, the empirical Bayes estimates will also be misspecified: The estimates of the \( \gamma \) parameters will be biased and the empirical Bayes shrinkage will lead to distortion in each group’s estimated equation. We shall return to this concern in Chapter 5 when we consider the problem of estimating the effectiveness of individual organizations.

Summary of Terms Introduced in This Chapter

Plausible value range for \( \beta_{q} \): We can compute a range of plausible values for any random level-1 coefficient, \( \beta_{q} \). The 95% plausible value range for \( \beta_{q} \) is \( \gamma_{p} \pm 1.96(\tau_{pq})^{1/2} \). Based on the assumption that the random effects at level 2 are normally distributed, we would expect to find values of \( \beta_{q,l} \) within range for 95% of the level-2 units.

Proportion of variance explained at level 1: An index of the proportion reduction in variance or “variance explained” at level 1 as \( X \) predictors are entered into the level-1 model. This is computed by comparing the residual \( \sigma_{r}^{2} \) estimate from a fitted model with the \( \sigma_{r}^{2} \) estimate from some “base” or reference model. The reference model chosen for computing these statistics at level 1 is often the one-way random-effects ANOVA model. (See Equation 4.20.)

Proportion of variance explained at level 2 in each \( \beta_{q} \): An index of the proportion reduction in variance or “variance explained” in each random level-1 coefficient (intercepts and slopes) as \( W \) predictors are added to the level-2 model for any \( \beta_{q,l} \). This is computed by comparing the residual \( \tau_{pq} \) estimate from a fitted model with the \( \tau_{pq} \) estimate from some “base” or reference model. The reference model chosen for computing these statistics at level 2 is often the random-coefficient regression model. (See Equation 4.24.)

Note

5 Applications in Organizational Research

- Background Issues in Research on Organizational Effects
- Formulating Models
- Case 1: Modeling the Common Effects of Organizations via Random-Intercept Models
- Case 2: Explaining the Differentiating Effects of Organizations via Intercepts- and Slopes-as-Outcomes Models
- Applications with Heterogeneous Level-1 Variance
- Centering Level-1 Predictors in Organizational Effects Applications
- Use of Proportion Reduction in Variance Statistics
- Estimating the Effects of Individual Organizations
- Power Considerations in Designing Two-Level Organization Effects Studies

Background Issues in Research on Organizational Effects

A number of conceptual and technical difficulties have plagued past analyses of multilevel data in organizational research. Among the most commonly experienced difficulties have been aggregation bias, misestimated standard errors, and heterogeneity of regression.

Aggregation bias can occur when a variable takes on different forms at different levels and therefore may have different effects at different organizational levels. In educational research, for example, the average social class of a school may have an effect on student achievement above and beyond the effect of the individual child’s social class. At the student level, social class provides a measure of the intellectual and tangible resources in a child’s home environment. At the school level, it is a proxy measure of a school’s
resources and normative environment. Hierarchical linear models help resolve this confounding by facilitating a decomposition of any observed relationship between variables, such as achievement and social class, into separate level-1 and level-2 components.

Misestimated standard errors occur with multilevel data when we fail to take into account the dependence among individual responses within the same organization. This dependence may arise because of shared experiences within the organization or because of the ways in which individuals were initially drawn into the organization. Hierarchical linear models resolve this problem by incorporating into the statistical model a unique random effect for each organizational unit. The variability in these random effects is taken into account in estimating standard errors. In the terminology of survey research, these standard error estimates adjust for the intraclass correlation (or related to it, the design effect) that occurs as a result of cluster sampling.

Heterogeneity of regression occurs when the relationships between individual characteristics and outcomes vary across organizations. Although this phenomenon has often been viewed as a methodological nuisance, the causes of heterogeneity of regression are often of substantive interest. Hierarchical linear models enable the investigator to estimate a separate set of regression coefficients for each organizational unit, and then to model variation among the organizations in their sets of coefficients as multivariate outcomes to be explained by organizational factors. Burstein (1980) provided an early review of this idea of slopes-as-outcomes.

Formulating Models

Many questions about how organizations affect the individuals within them can be formulated as two-level hierarchical linear models. At level 1, the units are persons and each person’s outcome is represented as a function of a set of individual characteristics. At level 2, the units are organizations. The regression coefficients in the level-1 model for each organization are conceived as outcome variables that are hypothesized to depend on specific organizational characteristics.

Person-Level Model (Level 1)

We denote the outcome for person \( i \) in organization \( j \) as \( Y_{ij} \). This outcome is represented as a function of individual characteristics, \( X_{qij} \), and a model error \( r_{ij} \):

\[
Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \cdots + \beta_{Qj}X_{Qij} + r_{ij}, \tag{5.1}
\]

where we will initially assume that \( r_{ij} \sim N(0, \sigma^2) \). (Extensions to models with heterogeneous level-1 variances are introduced in a later section of this chapter.)

The regression coefficients \( \beta_{qj}, q = 0, \ldots, Q \), indicate how the outcome is distributed in organization \( j \) as a function of the measured person characteristics. We therefore term these coefficients distributive effects.

Organization-Level Model (Level 2)

The effects for each organization, captured in the set of \( \beta_{qj} \)'s in Equation 5.1, are presumed to vary across units. This variation is in turn modeled in a set of \( Q + 1 \) level-2 equations—one for each of the regression coefficients from the level-1 model. Each \( \beta_{qj} \) is conceived as an outcome variable that depends on a set of organization-level variables, \( W_{sij} \), and a unique organization effect, \( u_{qj} \). Each \( \beta_{qj} \) has a model of the form

\[
\beta_{qj} = \gamma_{q0} + \gamma_{q1}W_{1j} + \gamma_{q2}W_{2j} + \cdots + \gamma_{qS}W_{S_{qj}} + u_{qj}. \tag{5.2}
\]

distributive effects in organization \( j \) = unique effect associated with organization \( j \)

distribution of outcomes within organization \( j \) = characteristics on the organizational level

where a unique set of predictors \( W_{s} (s = 1, \ldots, S_{q}) \) may be specified for each \( \beta_{qj} \).

The \( \gamma_{q} \) coefficients capture the influence of organizational variables, \( W_{sij} \), on the within-organization relationships represented by \( \beta_{qj} \). We assume that the set of \( Q + 1 \) level-2 random effects is multivariate normally distributed. Each \( u_{qj} \) has a mean of 0, some variance, \( \tau_{q0} \), and with covariances, \( \tau_{qq'} \), between any two random effects \( q \) and \( q' \). These are the standard level-2 model assumptions introduced in Chapter 3 and discussed in more detail in Chapter 9.

The next two sections of this chapter demonstrate how this model can be applied to investigate two broad classes of organizational effects. In “Case 1,” some aspect of the organization such as its technology, structure, or climate exerts a common influence on each person within it. Such organization effects modify only the mean level of the outcome for the organization. They leave unchanged the distribution of effects among persons within the organization. In statistical terms, only the intercept, \( \beta_{0j} \), varies across organizations; all other level-1 coefficients remain constant. As discussed in this section, such problems involve the use of random-intercept models.
In "Case 2," organization effects may modify not only the mean level of outcomes but also how effects are distributed among individuals. In statistical terms, the intercept and regression slopes vary among the units. This section details this application of the full hierarchical linear model with both intercepts- and slopes-as-outcomes.

The remainder of the chapter consists of a series of "special topics" concerning the design and use of hierarchical models in organizational applications. Much of the material presented here is new to the second edition. We demonstrate how the basic hierarchical model can be generalized to model heterogeneous variances at level 1. Following this, we detail how the choice of centering for level-1 variables affects the estimation of random level-1 coefficients, $\beta_j$, fixed effects, $\gamma$, and the variance-covariance components $\tau_{00}$, $\tau_{02}$, $\sigma^2$. We then discuss some complications that can arise in interpreting proportion in variance-explained statistics in more complex organizational effects models. Following this, we describe how to use the empirical Bayes estimates of level-1 coefficients as performance indicators for specific organizational units and discuss validity concerns that may arise in such applications. The chapter concludes with an introduction to power considerations in designing new data collections for two-level organizational effects studies.

Case 1: Modeling the Common Effects of Organizations via Random-Intercept Models

The basic problem in this case is that the key predictors are measured at the organization level, but the outcome variable is measured at the person level. Historically, such data have raised questions about the appropriate unit of analysis (organization or person) and the problems associated with either choice. If the data are analyzed at the person level, thereby ignoring the nesting of individuals within organizational units, the estimated standard errors will be too small, and the risk of type I errors inflated. Alternatively, if the data are analyzed at the organization level, using the means of the person responses as the outcome, it becomes problematic to incorporate other level-1 predictors into the analysis. In addition, inefficient and biased estimates of organizational effects can result. The key fact is that random variation and structural effects may exist at both levels, and a multilevel modeling framework is required to explicitly represent these features.

A Simple Random-Intercept Model

The basic idea of the random-intercept model was previously introduced in Chapter 2. The key feature of such models is that only the intercept parameter

in the level-1 model, $\beta_{0j}$, in Equation 5.1, is assumed to vary at level 2. Specifically, the organization model at level 2 consists of

$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_{1j} + \gamma_{02}w_{2j} + \cdots + \gamma_{08}w_{8j} + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

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

$$\cdots$$

$$\beta_{qj} = \gamma_{q0}.$$  

[5.3]

Example: Examining School Effects on Teacher Efficacy

Bryk and Driscoll (1988) used the Administrator and Teacher Supplement of the High School and Beyond Survey to investigate how characteristics of school organization were related to teachers' sense of efficacy in their work. Specifically, they hypothesized that teachers' efficacy would be higher in schools with a communal, rather than a bureaucratic, organizational form. The data consisted of responses from over 8,000 teachers nested within 357 schools. The average school sample size was 22 teachers.

The teacher-level model specified that the sense of efficacy varied among teachers within a school. Because no teacher variables were considered as level-1 predictors, Equation 5.1 reduced to

$$Y_{ij} = \beta_{0j} + r_{ij},$$  

[5.4]

where $Y_{ij}$ was the reported efficacy for teacher $i$ in school $j$, and $\beta_{0j}$ was the true mean level of efficacy in school $j$.

Three school-level models were estimated. The first was an unconditional model for $\beta_{0j}$. This results in a one-way random-effects ANOVA model, which partitions the total variance in $Y_{ij}$ into its within- and between-school components. These estimates of $\sigma^2$ and $\tau_{00}$, respectively, prove helpful in evaluating the results of subsequent models. The second model examined the effects of several measures of school composition and size on teacher efficacy (see Table 5.1 for description of variables). Formally,

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \text{(MEAN BACKGROUND)}_j + \gamma_{02} \text{(MEAN SES)}_j$$

$$+ \gamma_{03} \text{(HI MINORITY)}_j + \gamma_{04} \text{(SIZE)}_j + \gamma_{05} \text{(ETHNIC MIX)}_j$$

$$+ \gamma_{06} \text{(SES MIX)}_j + u_{0j}.$$  

[5.5]
TABLE 5.1 Description of Variables Used in the Study of the Effects of School Organization on Teacher Efficacy

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEACHERS EFFICACY</td>
<td>A factor composite of five teacher responses about their sense of satisfaction in their work. It is a standardized measure (i.e., mean = 0; sd = 1.0).</td>
</tr>
<tr>
<td>MEAN BACKGROUND</td>
<td>A factor composite of four items about students' academic experiences prior to high school (e.g., retained in grade) and initial placement (e.g., remedial English or math). It is a standardized variable (mean = 0; sd = 1.0) with positive scores indicating a stronger background.</td>
</tr>
<tr>
<td>MEAN SES</td>
<td>Average social class of students in the school. It is a standardized variable with positive values indicating more affluent schools.</td>
</tr>
<tr>
<td>HI MINORITY</td>
<td>A dummy variable indicating schools with minority enrollment in excess of 40%.</td>
</tr>
<tr>
<td>SIZE</td>
<td>The natural log of number of students enrolled in the school.</td>
</tr>
<tr>
<td>ETHNIC MIX</td>
<td>A standardized measure of the diversity in students' ethnicity within the school. Low values imply a single ethnic group. High positive values indicate significant student representation in several ethnic groups.</td>
</tr>
<tr>
<td>SES MIX</td>
<td>Like ethnic mix, this is a standardized measure of the social class diversity within the school. Positive values indicate a socially heterogeneous school.</td>
</tr>
<tr>
<td>COMMUNAL</td>
<td>A composite measure based on 23 separate indices of the extent to which a school has a communal organization. It is a standardized measure with positive values indicating a greater frequency of shared activities, consensus, common beliefs, teacher collegiality, and a broader teacher role. Low values indicate a more segregated, specialized, bureaucratic organization.</td>
</tr>
</tbody>
</table>

NOTE: For a further discussion of the measures, see Bryk and Driscoll (1988).

The third model added a measure of the degree of communal organization in the school, COMMUNAL, to Equation 5.5. In both models, the residual school-specific effects, \( \mu_0 \), were assumed normally distributed with mean 0 and variance \( \tau_{00} \).

One-Way Random-Effects ANOVA Model. The analysis began with fitting an unconditional model for \( \beta_0 \) at level 2. The estimate for the within-school or level-1 variance [i.e., \( \text{Var}(r_{ij}) = \sigma^2 \)] was 0.915. This estimate remains the same for the three analyses discussed here, because the level-1 model (Equation 5.4) is identical for all three. The overall variability among the true school means on teacher efficacy [i.e., \( \text{Var}(\beta_0) = \tau_{00} \)] was 0.084. This resulted in an intraclass correlation of 0.084 (see Equation 4.6), and an estimated reliability for the school means on teacher efficacy of 0.669 (see Equations 4.7 and 4.8).

Two Explanatory Models at Level 2. The first two columns in Table 5.2 present the results of the hierarchical linear model analyses. In the top panel are the estimates for the compositional model and in the bottom panel are the results after the COMMUNAL variable was added. In the compositional model (top panel) students' MEAN BACKGROUND (\( \gamma_{01} = 0.044, \text{se} = 0.020 \)) and the school's MEAN SES (\( \gamma_{02} = 0.133, \text{se} = 0.023 \)) were positively related to teachers' sense of efficacy. School SIZE had a significant

<table>
<thead>
<tr>
<th>TABLE 5.2 Effects of School Organization on Teacher Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical Analysis</td>
</tr>
<tr>
<td>Coefficient</td>
</tr>
<tr>
<td>Compositional model</td>
</tr>
<tr>
<td>MEAN BACKGROUND, ( \gamma_{01} )</td>
</tr>
<tr>
<td>MEAN SES, ( \gamma_{02} )</td>
</tr>
<tr>
<td>HI MINORITY, ( \gamma_{03} )</td>
</tr>
<tr>
<td>SIZE, ( \gamma_{04} )</td>
</tr>
<tr>
<td>ETHNIC MIX, ( \gamma_{05} )</td>
</tr>
<tr>
<td>SES MIX, ( \gamma_{06} )</td>
</tr>
<tr>
<td>Proportion of variance accounted for</td>
</tr>
</tbody>
</table>

Communal model

| Coefficient | se | Coefficient | se | Coefficient | se |
| Compositional model   |               |               |               |               |
| MEAN BACKGROUND, \( \gamma_{01} \) | 0.038 | 0.017 | 0.040 | 0.013 | 0.033 | 0.018 |
| MEAN SES, \( \gamma_{02} \) | 0.015 | 0.022 | 0.015 | 0.017 | 0.019 | 0.023 |
| HI MINORITY, \( \gamma_{03} \) | -0.055 | 0.040 | -0.056 | 0.031 | -0.051 | 0.041 |
| SIZE, \( \gamma_{04} \) | 0.061 | 0.026 | 0.062 | 0.021 | 0.060 | 0.025 |
| ETHNIC MIX, \( \gamma_{05} \) | -0.014 | 0.016 | -0.014 | 0.013 | -0.014 | 0.017 |
| SES MIX, \( \gamma_{06} \) | 0.001 | 0.020 | 0.002 | 0.016 | -0.000 | 0.020 |
| COMMUNAL, \( \gamma_{07} \) | 0.504 | 0.045 | 0.507 | 0.035 | 0.493 | 0.045 |
| Proportion of variance accounted for | 0.631 | 0.054 | 0.426 |
| Incremental variance | 0.286 | 0.025 | 0.192 |

NOTE: Residual variance estimates for hierarchical analyses: \( \tau_{00} \) (compositional model) = 0.055; \( \tau_{00} \) (communal model) = 0.031.
negative effect \( \hat{\gamma}_4 = -0.066, \text{ se} = 0.027 \). The effects of the other three level-2 predictors were small both in absolute terms and in comparison to their estimated standard errors.

The results for the communal organization model were quite startling. The estimated effect of COMMUNAL \( \hat{\gamma}_3 = 0.504, \text{ se} = 0.045 \) was by far the largest—almost an order of magnitude bigger than all the others. (Note: This interpretation depends on the fact that all of the level-2 predictors, except the dummy variable HI MINORITY, were standardized to mean = 0, sd = 1.0.) This means that teacher efficacy was substantially higher in schools with a communal organization, even after controlling for compositional differences among schools. We also note that the effects of MEAN SES found in the composition model have largely disappeared (0.015 versus 0.133). This suggests that the positive levels of teacher efficacy found in high social-class schools may reflect the greater prevalence of communal organizational features in these schools. Notice also that the school-size effect has become positive, implying somewhat higher levels of teacher efficacy in larger schools after controlling for communal organization. As Bryk and Driscoll (1988) explain, teacher efficacy tends to be lower in large schools because these schools are less likely to be communal. Once this effect is controlled for, however, large schools appear to promote somewhat greater efficacy presumably by virtue of the greater resources and expanded professional opportunities typically found there.

After controlling for composition effects, as specified in Equation 5.5, the residual variability at level 2, \( \hat{\tau}_{01} \), was 0.055. We refer to this as \( \hat{\tau}_{00}(\text{compositional model}) \). The model accounted for 34.5% of the total parameter variance among schools in mean levels of teacher efficacy. This proportion reduction in variance statistic was computed using the procedure introduced in Chapter 4. Specifically,

\[
\text{Proportion variance explained} = \frac{\hat{\tau}_{00}(\text{unconditional}) - \hat{\tau}_{00}(\text{compositional model})}{\hat{\tau}_{00}(\text{unconditional})} = \frac{0.084 - 0.055}{0.084} = 0.345,
\]

where \( \hat{\tau}_{00}(\text{unconditional}) \) is the overall variability in the true school means as estimated from the one-way random-effects ANOVA model.

The percentage of variance explained jumped to 63.1% after COMMUNAL was entered into the model. That is, the residual variance in \( \hat{\beta}_{0j}, \hat{\tau}_{00} \) (communal model), was 0.031. Replacing \( \hat{\tau}_{00}(\text{compositional model}) \) in

Equation 5.6 with \( \hat{\tau}_{00}(\text{communal model}) \) yields a proportion of variance explained as

\[
\frac{0.084 - 0.031}{0.084} = 0.631.
\]

The incremental variance explained by adding COMMUNAL to the model was 28.6%. This statistic is just the difference between the proportion reduction-in-variance statistics calculated in Equations 5.6 and 5.7.

Comparison of Results with Conventional Teacher-Level and School-Level Analyses

Table 5.2 also presents results from teacher (level-1) and school (level-2) analyses. A comparison of these conventional alternatives to the multilevel results helps to clarify some of the basic features of estimates in random-intercept models. We compare below the fixed-effect estimates, the standard errors of these estimates, and the variance-explained statistics typically reported with such analyses.

**Fixed Effects.** Notice in Table 5.2 that the estimates for the regression coefficients are quite similar across the three analyses. The hierarchical estimates, however, are somewhat closer to the results from the teacher analysis, which will generally be the case.

As presented in Chapter 3, the estimators for the level-2 coefficients in a hierarchical linear model can be viewed as weighted least squares estimators where the weights are of the form

\[
\Delta_j^{-1} = (V_j + \tau_{00})^{-1}.
\]

Assuming a homogeneous level-1 variance (i.e., \( \sigma_j^2 = \sigma^2 \) for all \( J \)), then \( V_j = \sigma^2 / n_j \) and the variation in the weights depends strictly on \( n_j \).

In comparison, the typical OLS level-1 analysis of these data is also weighted, but the weights are just \( n_j \). Specifically, suppose we estimate a simple univariate model where the teacher outcome, \( Y_{ij} \), depends on one school characteristic, \( W_j \):

\[
Y_{ij} = \gamma_0 + \gamma_1 W_j + e_{ij}.
\]

The estimator for \( \gamma_1 \) based on an OLS level-1 analysis is simply

\[
\hat{\gamma}_1 = \frac{\sum_{1} \sum_{j} n_j (W_j - \bar{W})(Y_{ij} - \bar{Y}_i)}{\sum_{j} n_j (W_j - \bar{W})^2}.
\]
Hierarchical Linear Models

Notice that the numerator and denominator are sums of squares and cross-products, weighted by \( n_j \) instead of \( \Delta_j^{-1} \).

In contrast, the level-2 analysis is unweighted. The comparable univariate model would be

\[
Y_{ij} = \gamma_0 + \gamma_1 W_j + \gamma_{ij} \tag{5.11}
\]

and the corresponding estimator for \( \gamma_1 \) is

\[
\hat{\gamma}_1 = \frac{\sum_j(W_j - \bar{W})(\bar{Y}_j - \bar{Y})}{\sum_j(W_j - \bar{W})^2}, \tag{5.12}
\]

where

\[
\bar{Y} = \frac{\sum Y_{ij}}{J} \quad \text{and} \quad \bar{W} = \frac{\sum W_j}{J}.
\]

All three estimators are unbiased, but the hierarchical estimator is the most efficient. Variations in results among the three analyses will depend on the degree of imbalance in the \( n_j \). (If the sample sizes \( n_j \) are identical for each of the \( J \) organizations, the three estimators are the same.) In the Bryk and Driscoll (1988) application, the \( n_j \) were not grossly different. Most schools had between 20 and 30 cases. Thus, the level-2 results were quite similar in this application.

In general, a robustness concern arises when estimating fixed effects with a level-2 analysis in the presence of unbalanced data. A unit with a very small sample size can easily become an outlier or leverage point because of the instability associated with the limited amount of information about that unit. The weighting employed in the hierarchical and level-1 analyses protect against this.

Standard Errors of the Fixed Effects. As noted in the introduction to this chapter, the standard errors produced by a level-1 analysis will generally be too small, because this analysis fails to take into account the fact that level-1 units are not independent but rather are actually clustered within level-2 units. In the compositional analysis, for example, the level-1 standard-error estimates are about a third smaller than those provided by the hierarchical and level-2 analyses.

A direct comparison of the formulas for the three different standard errors is difficult when sample sizes are unequal. Some basic features can be ascertained, however, if we consider the balanced data case with a single \( W_j \) predictor. It can readily be shown that the expected values of the estimators for the sampling variance for \( \gamma_1 \) from the hierarchical linear model and level-2 analyses are identical:

\[
E[\text{Var}(\hat{\gamma}_1)]_{\text{hierarchical}} = \frac{V + \tau_{00}}{\sum_j(W_j - \bar{W})^2}.
\]

As for the level-1 analysis, the expected value of the sampling variance estimator is

\[
E[\text{Var}(\hat{\gamma}_1)]_{\text{level-1}} = \frac{J(n - 1)\sigma^2 + (J - 2)n(V + \tau_{00})}{(Jn - 2n) \sum_j(W_j - \bar{W})^2}.
\]

When the level-1 and level-2 samples (\( n \) and \( J \), respectively) are large, the ratio of the expected sampling variance from a level-1 analysis, Equation 5.14, to that from a hierarchical analysis (or, equivalently, a level-2 analysis), Equation 5.13, is approximately

\[
\frac{E[\text{Var}(\hat{\gamma}_1)]_{\text{level-1}}}{E[\text{Var}(\hat{\gamma}_1)]_{\text{hierarchical}}} \approx 1 - \lambda,
\]

where \( \lambda = \tau_{00}/[(\sigma^2/n) + \tau_{00}] \), and is the reliability of the OLS estimated school means, \( \beta_{0j} \).

Equation 5.15 closely approximates the empirical results reported in Table 5.2. With an average level-1 sample size of 22 teachers per school, a level-1 variance estimate, \( \hat{\sigma}^2 = 0.915 \), and an estimate of \( \tau_{00} \) for the compositional model of 0.055,

\[
1 - \lambda = 1 - \frac{0.055}{(0.915/22) + (0.055)} = 0.431.
\]

As for the relative size of the standard errors, their ratio is simply \((1 - \lambda)^{1/2}\), which, for the compositional model, yields a value of 0.657. This ratio closely corresponds to the results reported in the top panel of Table 5.2. Visual inspection indicates that the level-1 standard errors are approximately two thirds of the more appropriate values reported by the hierarchical and level-2 analyses.

In sum, the hierarchical analysis captures the best features of both the level-1 and level-2 analyses. It provides unbiased and efficient estimates of the fixed effects, which are more closely approximated by the level-1 analysis, and provides proper standard error estimates, regardless of the degree of within-unit clustering, that are more closely approximated by the level-2 analysis.

The results obtained above are typical of what one might routinely encounter in this type of analysis. The fixed-effects estimates will often be \( \gamma \) similar across the three analyses; however, the estimated standard errors will not.

Variance-Explained Statistics. The estimates of the proportion of variance explained from a hierarchical analysis may be quite different from those generated in conventional level-1 or level-2 analyses and may lead to different
conclusions. In the Bryk and Driscoll (1988) study, for example, a judgment about the importance of communal organization depended considerably on the analysis considered (bottom panel in Table 5.2). The incremental reduction in variance associated with COMMUNAL was 28.6% in the hierarchical linear model analysis. The comparable statistics from the teacher and school-level analyses were 2.5% and 19.2% respectively. Although most analyses would probably judge 28.6% sufficiently substantial to merit further consideration of the construct under study, the 2.5% statistic could easily lead to the opposite inference.

To understand why the proportion of variance accounted for by COMMUNAL is so different requires a closer consideration of how the total outcome variability is partitioned in the three analyses. In a random-intercept problem, level-2 variables such as COMMUNAL can only account for variation among the true school means, \( \beta_{0j} \). That is, only the parameter variation, \( \tau_{00} \), is explainable. (This is why we use \( \tau_{00} \) from the unconditional model as the denominator in the proportion reduction in variance statistic.) The 28.6% variance explained by COMMUNAL implies that a substantial portion of the variation among the true school means is associated with variation in school organization. Relative to all other school-level sources of variation, communal organization is indeed important.

In comparison, the level-1 analysis employs the total outcome variability in \( Y_{ij} \), \( \tau_{00} + \sigma^2 \), as the denominator for the variance-explained statistics. The within-unit variation, \( \sigma^2 \), however, reflects individual effects and errors of measurement in the outcome variable, both of which are unexplainable by organizational features. Judged against this standard, some researchers might erroneously conclude that the COMMUNAL is trivially small.

In general, the relative variance explained by a hierarchical versus a level-1 analysis depends on the ratio

\[
\frac{\text{Variance explained (level 1)}}{\text{Variance explained (hierarchical)}} \approx \frac{\tau_{00}}{\tau_{00} + \sigma^2} = \rho, \tag{5.16}
\]

where \( \rho \) is the intraclass correlation coefficient (see Equation 2.10). Note that the intraclass correlation represents the theoretically maximal amount of the total variance in the outcome \( Y_{ij} \) that is explainable by all school factors. As noted earlier, the estimated intraclass correlation was 0.084 for the teacher-efficacy data. We can use \( \hat{\rho} \) to relate the variance-explained statistics from the level-1 and hierarchical analyses. For example,

\[
\hat{\rho} \times \left[ \frac{\text{Incremental variance explained (hierarchical)}}{\text{Incremental variance explained (level 1)}} \right] \\
0.084 \times [0.286] \approx 0.024.
\]

A similar formula can be derived for comparing the variance-explained statistics from the hierarchical and level-2 analyses. The denominator for the variance-explained statistic in a level-2 analyses is \( \tau_{00} + \sigma^2/n_j \), which is just the total variance of the sample means. Thus, the relative variance explained by a hierarchical versus level-2 analysis is approximately

\[
\frac{\text{Variance explained (level 2)}}{\text{Variance explained (hierarchical)}} \approx \frac{\tau_{00}}{\tau_{00} + (\sigma^2/n_j)} = \lambda, \tag{5.17}
\]

where \( \lambda \) is the average reliability of the \( \bar{Y}_{ij} \), as estimates of \( \mu_j \) (see Equation 4.8) based on an average level-1 sample size of \( \bar{n}_i \). For the teacher-efficacy data, \( \lambda \) was 0.669, and the variances explained by the level-2 analyses were approximately two thirds of those represented in the corresponding hierarchical analyses.

In sum, the variance-explained statistic from the hierarchical analysis provides the clearest evidence for making judgments about the importance of level-2 predictors. They are not affected by the degree of clustering as the level-1 statistics are (i.e., the dependence on \( \rho \)), nor are they affected by the unreliability of \( \bar{Y}_{ij} \) as the level-2 statistics are. Further, because good estimates of \( \lambda \) and \( \rho \) are not generally available with conventional analyses, the analyst has no way to assess the explanatory power of a set of level-2 predictors relative to the maximum amount explainable by any model. Intuitively, this is what variance-explained statistics should tell us.

A Random-Intercept Model with Level-1 Covariates

In the previous example, we estimated the relationship between organizational characteristics and mean outcomes. We made no attempt to adjust the level-2 effect estimates for the different characteristics of the individuals in the various organizations.

In general, statistical adjustments for individual background are important for two reasons. First, because persons are not usually assigned at random to organizations, failure to control for background may bias the estimates of organization effects. Second, if these level-1 predictors (or covariates) are strongly related to the outcome of interest, controlling for them will increase the precision of any estimates of organizational effects and the power of hypothesis tests by reducing unexplained level-1 error variance, \( \sigma^2 \).

The formal model for this type of analysis was introduced in Chapter 2. At level 1,

\[
Y_{ij} = \beta_{0ij} + \beta_1(X_{1ij} - \bar{X}_{1..}) + \beta_2(X_{2ij} - \bar{X}_{2..}) \\
+ \cdots + \beta_q(X_{qij} - \bar{X}_{q..}) + r_{ij} \tag{5.18}
\]

\]
The level-2 model is Equation 5.3. Because each covariate is centered around its respective grand mean, the random intercept, $\beta_0j$, is now the adjusted mean rather than the raw mean. As in ANCOVA models, Equation 5.18 assumes homogeneous level-1 coefficients for $\beta_{1j}, \ldots, \beta_{0j}$. The validity of this assumption can be easily tested using the methods described in Chapter 3. If needed, any level-1 coefficient can be specified as either non-randomly varying or as a random effect.

**Example: Evaluating Program Effects on Writing**

This example uses data from the Cognitive Strategies in Writing Project (Englert et al., 1988). The project sought to improve children's writing and to enhance children's self-perceptions of academic competence through a variety of strategies. The outcome variable was a measure of perceived academic self-competence (mean $= 2.918$; sd $= 0.580$) for which a pretest, denoted $X_{ij}$, served as the covariate. The study involved 256 children in 22 classrooms in a standard two-group design, with 15 experimental classrooms and 7 control classrooms. Because classroom teachers implemented the treatments to intact classrooms, we have, in classical terms, a nested or hierarchical design: Students are nested within classrooms with the treatment administered at the classroom level. As in the previous example, we first present the results of a hierarchical analysis. We then compare these results to conventional alternatives: an ANCOVA at the student level ignoring classes and an ANCOVA based on class means.

For the hierarchical analysis, the level-1 model was

$$ Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}) + r_{ij}, \quad [5.19] $$

where

$\gamma_{0j}$ is the self-perceived competence of child $i$ in class $j$ ($j = 1, \ldots, 22$ classrooms);

$\beta_{0j}$ is the adjusted mean outcome in class $j$ after controlling for differences in pretest status; and

$\beta_{1j}$ is the fixed level-1 covariate effect.

A preliminary analysis specified a model where both $\beta_{0j}$ and $\beta_{1j}$ were random and tested the homogeneity hypothesis for the covariate effect $[H_0: \text{Var}(\beta_{1j}) = 0]$. Because the null hypothesis was retained for $\beta_{1j}$, it was appropriate to assume a fixed effect for the pretest covariate. The final level-2 model was

$$ \beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}, \quad \beta_{1j} = \gamma_{10}, \quad [5.20] $$

where

$W_j$ is a treatment-indicator variable (1 = experimental; 0 = control);

$\gamma_{00}$ is the adjusted mean achievement in the control-group classrooms;

$\gamma_{01}$ is the treatment effect; and

$\gamma_{10}$ is the pooled within-classroom regression coefficient for the level-1 covariate.

The results of this analysis appear in Table 5.3. The estimated difference between experimental and control means adjusted for the pretest was $1.188$ ($t = 1.87$, df $= 20$, $p<$ .04). The estimated pooled within-classroom regression slope for the posttest on the pretest was $0.396$ ($t = 7.02$, $p < .001$).

**Comparison of Results with Conventional Student- and Classroom-Level Analyses**

The model for the student-level analyses was

$$ Y_i = \gamma_{00} + \gamma_{01}W_i + \gamma_{10}(X_i - \bar{X}) + r_i, \quad [5.21] $$

where the parameters $\gamma_{00}$, $\gamma_{01}$, and $\gamma_{10}$ represent the intercept, the treatment effect, and the covariate effect, respectively, and $\bar{X}$ is the mean pretest score (i.e., $\sum_{i=1}^{N} X_i / N$). Notice that the $j$ subscript has disappeared because class membership is ignored as is the effect associated with classrooms (i.e., $u_{0j}$) in Equation 5.20.

The results for this analysis appear in the second column of Table 5.3. They indicate that experimental children developed a significantly higher perceived self-competence than did the control children ($\hat{\gamma}_{00} = 0.160$, $t = 2.17$, $p<$ .02). The estimated pooled within-treatment groups regression slope for the posttest on the pretest was $0.406$ ($t = 7.25$, $p < .001$).

**TABLE 5.3 Effects of Experimental Instruction on Self-Perceived Competence in Writing**

<table>
<thead>
<tr>
<th></th>
<th>Hierarchical Analysis$^a$</th>
<th>Student-Level Analysis$^b$</th>
<th>Classroom-Level Analysis$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient $\gamma_{00}$</td>
<td>$se$</td>
<td>Coefficient $\gamma_{00}$</td>
</tr>
<tr>
<td>Intercept, $\gamma_{00}$</td>
<td>2.774</td>
<td>0.084</td>
<td>2.802</td>
</tr>
<tr>
<td>Treatment indicator, $\gamma_{01}$</td>
<td>0.188</td>
<td>0.100</td>
<td>0.160</td>
</tr>
<tr>
<td>Pretest, $\gamma_{10}$</td>
<td>0.396</td>
<td>0.056</td>
<td>0.406</td>
</tr>
</tbody>
</table>

$^a$ Residual Variance estimate: $\hat{\sigma}^2 = 0.258$.

$^b$ Estimated residual variance = 0.273.

$^c$ Estimated residual variance = 0.087.
As for the classroom-level analyses, the model was
\[
\bar{Y}_i = \gamma_0 + \gamma_1 W_i + \gamma_0 (\bar{X}_{-j} - \bar{X}_c) + u_{0i},
\]  
where \( j = 1, \ldots, 22 \) classrooms. Here \( \bar{X}_{-j} \) is the pretest mean for class \( j \); \( \bar{Y}_i \) is the posttest mean for class \( j \). The grand mean for the pretest, \( \bar{X}_c \), is the mean of the classroom means (i.e., \( \bar{X}_c = \sum_{j=1}^{22} \bar{X}_{-j} / J \)). These results appear in the third column of Table 5.3. Although the estimated treatment-effect size (\( \hat{\gamma}_0 = 0.209 \)) is actually larger than the hierarchical and classroom-level estimates, it is not statistically significant due to the substantially larger standard error [\( \text{se}(\hat{\gamma}_0) = 0.135 \)]. The estimated covariate effect, \( \hat{\gamma}_{10} = 0.649 \), is also substantially larger, as is its standard error of .223.

It may seem surprising that the hierarchical analysis produced inferences similar to those of the student-level analysis and different from the class-level analysis. In each case, the test statistic for the effect of innovative instruction depends on a ratio of two quantities: the fixed-effect size estimate and the standard error of this estimate. A comparison of each is offered below.

Fixed Effects. The estimate for the treatment effect is reasonably similar in all three analyses, with the student-level analysis producing the smallest effect (.160), the class-level analysis the largest (.209), and the hierarchical estimate falling in between (.188). In all three analyses, the treatment-effect estimator is of the general form
\[
\hat{\gamma}_{10} = \hat{\mu}_{Y_i} - \hat{\mu}_{Y_c} - \hat{\beta}_{Y,X}(\hat{\mu}_{X_i} - \hat{\mu}_{X_c}).
\]  
where \( \hat{\mu}_{Y_i} \) and \( \hat{\mu}_{Y_c} \) are estimates of the posttest means for experimentals and controls, respectively. The key differences among the analyses are in the way \( \hat{\beta}_{Y,X} \) and the pre- and posttest means are estimated. Table 5.4 presents the relevant statistics from each analysis.

### Table 5.4 Treatment Effect Estimates: Hierarchical, Student-Level, and Classroom-Level Analyses

<table>
<thead>
<tr>
<th></th>
<th>Hierarchical Analysis</th>
<th>Student-Level Analysis</th>
<th>Classroom-Level Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{Y}_i )</td>
<td>2.968</td>
<td>2.980</td>
<td>2.964</td>
</tr>
<tr>
<td>( \bar{Y}_c )</td>
<td>2.742</td>
<td>2.754</td>
<td>2.780</td>
</tr>
<tr>
<td>( \bar{X}_{-j} )</td>
<td>2.895</td>
<td>2.921</td>
<td>2.855</td>
</tr>
<tr>
<td>( \bar{X}_c )</td>
<td>2.797</td>
<td>2.799</td>
<td>2.803</td>
</tr>
<tr>
<td>( \hat{\beta}_{Y,X} )</td>
<td>0.396</td>
<td>0.406</td>
<td>0.649</td>
</tr>
<tr>
<td>( \bar{Y}<em>i - \bar{Y}<em>c - \hat{\beta}</em>{Y,X}(\bar{X}</em>{-j} - \bar{X}_c) )</td>
<td>0.188</td>
<td>0.160</td>
<td>0.209</td>
</tr>
</tbody>
</table>

For the student-level analysis, \( \hat{\beta}_{Y,X} \) is the regression of the posttest on the pretest, pooled within the treatment and control groups. The hierarchical analysis is similar, except \( \beta_{Y,X} \) is pooled within each of the 22 classrooms. In contrast, the class-level analysis regresses the posttest means for the 22 classes on their corresponding pretest means. As Table 5.4 shows, the \( \beta_{Y,X} \) are quite similar in the hierarchical (\( \hat{\beta}_{Y,X} = .396 \)) and the student-level analyses (\( \tilde{\beta}_{Y,X} = .406 \)). The classroom-level estimate, however, is quite discrepant (\( \hat{\beta}_{Y,X} = .649 \)).

The three alternative methods also employ different estimators of the pre- and posttest means for each treatment group. Consider, for example, the experimental posttest mean. The hierarchical-analysis estimator is a weighted average,
\[
\bar{Y}_{(\text{hierarchical})} = \frac{\sum_j \Delta_j^{-1} \bar{Y}_{j,E} / \sum_j \Delta_j}{\sum_j \Delta_j^{-1} / \sum_j},
\]  
where \( \bar{Y}_{j,E} \) is the mean of the \( j \)th classroom in the experimental group. The weight, \( \Delta_j^{-1} \), is the precision of the corresponding sample mean. Thus, the hierarchical estimates are precision-weighted averages.

The student-level estimators weight by sample sizes:
\[
\bar{Y}_{(\text{student level})} = \frac{\sum_j n_j \bar{Y}_{j,E} / \sum_j n_j}{\sum_j n_j / \sum_j},
\]  
where \( n_j \) is the sample size of the \( j \)th classroom in the experimental group.

In contrast, the classroom-level analysis uses an unweighted average:
\[
\bar{Y}_{(\text{classroom level})} = \frac{\sum_j \bar{Y}_{j,E}/J}{\sum_j \bar{Y}_{j,E}/J},
\]  
where \( J \) is the number of classrooms in the experimental group.

When the reliability of individual classroom means varies significantly, the classroom-level estimate is likely to be inaccurate because it will be strongly influenced by extreme classroom means, which may result from unreliability. This is not true of the student-level or hierarchical estimators. In fact, when the precisions \( \Delta_j^{-1} \) are known or estimated accurately, the hierarchical weighting scheme is optimal. The classroom-level estimate is defensible only if the sample means are equally reliable.

**Standard Errors.** The standard errors of the treatment effect estimates are also different across the three analyses. The value of 0.074 for the student-level analysis is clearly misleading because, as noted earlier, this analysis fails
to take into account the dependence among the observations within classrooms. In essence, the student-level analysis assumes more information is present than is actually the case (i.e., it assumes that each individual response within a classroom provides an additional independent piece of information). But why is the standard error estimated under the hierarchical model (0.100) smaller than the estimate based on the class-level analysis (0.135)?

Apart from sample size, the standard error of the difference between two treatment groups in an ANCOVA depends on three factors: (a) the unexplained variance in the outcome, (b) the precision of the estimated regression coefficient for the covariate, and (c) the magnitude of the difference between the groups on the covariate. The hierarchical analysis is generally more powerful than the class-level analysis because factors (a) and (b) work in its favor.

In terms of factor (a), the unexplained variance in the outcome is smaller in the hierarchical analysis. The variance of $\bar{Y}_{j}$, classroom $j$’s sample mean, is $\Delta_j = \tau_0 + \sigma^2/n_j$. In a class-level analysis, only $\tau_0$ is potentially explainable by the covariate. In the hierarchical analysis, both $\tau_0$ and $\sigma^2$ may be explained. The reduction in $\sigma^2$ can be substantial if the level-1 covariate is strongly related to the outcome within classrooms, yielding a potentially substantial advantage in power over the class-level analysis.

In terms of factor (b), the precision of the estimated covariate effect in the hierarchical analysis is greater than in the class-level analysis, because the hierarchical analysis uses all of the data to estimate the covariate effect. In contrast, the class-level analysis uses only information about the covariation between the class means on the pre- and posttest.

Specifically, the $se(\hat{\beta}_{y|x})$ for the hierarchical analysis will generally be smaller than $se(\hat{\beta}_{y|x})$ from the class-level analysis, as is demonstrated in Table 5.3. This is significant because the standard error of the treatment effect, $se(\hat{\gamma}_{y|x})$, depends on the $se(\hat{\beta}_{y|x})$, which is obvious from an inspection of Equation 5.23.

In sum, the hierarchical analysis offers several advantages in this application. First, it is an honest model. Rather than erroneously assuming independent responses within classes, the hierarchical model takes into account the dependence among responses within classrooms.

Second, it provides efficient estimates of treatment effects in unbalanced, nested designs. Traditionally, the class-level analysis has been recommended as the preferred alternative to the student-level analysis because of the untractability of the assumption of independent errors. Researchers have lamented, however, that such analyses, although perhaps more appropriate, have low power to detect effects, and for this reason, researchers have tended to ignore this advice. The key point is that researchers no longer have to make a choice between a clearly untenable model (i.e., student-level analysis) and an honest but low-power alternative (class-level analysis). The hierarchical linear model properly represents the sources of variation in nested designs and provides efficient parameter estimates.

Finally, the hierarchical model enables a test for homogeneity of regression and provides a sensible way to proceed, regardless of the outcome. In this application, the regression of posttest on pretest was homogeneous with regard to classrooms, so we treated the covariate as a fixed effect. However, if the regression coefficients had been found to vary across classrooms, we could have built a model to predict such variation. Any unexplained variation in this level-1 coefficient would then be incorporated into inference about treatments.

Case 2: Explaining the Differentiating Effects of Organizations via Intercepts- and Slopes-as-Outcomes Models

In the applications discussed above, organizational characteristics exercised a common influence on all individuals within the organization. The sole effect of the organizational variable under study was to shift the mean level of the outcomes, leaving the distribution of outcomes otherwise unaffected. In this section we consider situations where organizational features affect level-1 relationships, either amplifying or attenuating them. The corresponding statistical model for such phenomena is the full hierarchical model represented in Equations 5.1 and 5.2. The within-organization relations are represented by the regression coefficients in the level-1 model. The effects of organization variables on each of these relationships is represented in the corresponding level-2 model.

Difficulties Encountered in Past Efforts at Modeling Regression Slopes-as-Outcomes

The use of regression coefficients or slopes-as-outcomes is appealing because it extends substantially the kinds of questions that organizational research can examine. Unfortunately, a number of technical difficulties inhibited past use of models that incorporate slopes-as-outcomes.

First, as a general rule, regression coefficients have considerably greater sampling variability than do sample means. If the sample within a unit is small, the regression coefficients will be estimated with large error. The resultant unreliability in slopes weakens our power to detect relationships in the level-2 model. This imprecision is exacerbated when the dispersion in the level-1 predictors is constrained. For example, students tend to be more
homogeneous in social class within schools than they are in a true random sample. As a result, the sampling variability of the estimated within-schools SES-achievement slope is increased. The analysis can produce negative slope estimates for individual schools even when the structural parameter is clearly positive (see Figure 4.2). This is particularly problematic because such outliers can exert undue influence on the level-2 results.

Second, the sampling precision of the estimated slopes varies across units depending on the data-collection design used within each unit. But ordinary least squares, the estimation method typically used for the level-2 analysis, assumes equal variances across units on the dependent variable. Ignoring the variation in sampling precision across units results in a weakened efficiency in parameter estimation that further limits our ability to detect relationships between slopes and the level-2 variables hypothesized to account for them.

Third, the total variability in the estimated slopes consists of two components. First, there may be real differences across organizations in the slope parameters. It is essential, however, to distinguish between this parameter variance and the error variance in the slope estimates. This distinction becomes especially important when we attempt to interpret the results from the level-2 model. As noted earlier in this chapter, only parameter variance in the level-1 coefficients is potentially explainable by level-2 predictors. In many applications, much of the observed variance in the slopes is error variance for the reasons noted above. A level-2 model that explains only a small percentage of the observed variance in a regression slope might be discounted, when in fact it is explaining a very large portion of what can, in principle, be explained. Unfortunately, the simple slopes-as-outcomes model provides little guidance in this regard.

Fourth, to include multiple slopes-as-outcomes in the level-2 model requires us to take into account the special covariance structure that exists among the multiple-regression coefficients estimated for each level-2 unit. In the absence of such a model, further weakened precision is a likely result.

Fifth, in many applications, the sample of organization members will not support an ordinary least squares regression for every organization. If the sample size for a particular organization is very small, or if that sample does not vary for a particular $X$, it will not be possible to compute the regression. Using the slopes-as-outcomes approach, such organizations must be discarded, possibly biasing the sample as well as reducing precision. Such organizations need not be discarded when using a hierarchical model with maximum likelihood estimation.

### Example: The Social Distribution of Achievement in Public and Catholic High Schools

Lee and Bryk (1989) used hierarchical analyses on a subset data from the High School and Beyond Survey, similar to that used in Chapter 4, to examine whether academic achievement had a more equitable social distribution in the Catholic than in the public sector. Specifically, they drew a sample of 74 Catholic high schools and a random subsample of 86 public high schools. Data were combined from two cohorts of students to increase the level-1 sample sizes, $n_j$, to yield a total sample size, $N$, of 10,999 students. Table 5.5 describes selected variables used in their analyses. We discuss below some of their analyses and comment on the logic involved in using an intercepts- and slopes-as-outcomes model to explain the social distribution of achievement in Catholic and public high schools.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATH ACHIEVEMENT</td>
<td>A mathematics test in senior year (mean = 12.92, sd = 6.70).</td>
</tr>
<tr>
<td>SES</td>
<td>A composite measure of social class provided by High School and Beyond. For the analytic sample it had a mean of approximately zero and standard deviation of 0.8.</td>
</tr>
<tr>
<td>MINORITY</td>
<td>A dummy variable (1 = black or Hispanic; 0 = other).</td>
</tr>
<tr>
<td>BACKGROUND</td>
<td>A composite measure of students' academic background up to high school entry. It includes information on retention in elementary school, assignment to remedial classes in 9th grade, and educational expectation at high school entry. It is a standardized measure (mean = 0; sd = 1.0).</td>
</tr>
<tr>
<td>SECTOR</td>
<td>An effects-coded variable (1 = Catholic; -1 = public).</td>
</tr>
<tr>
<td>AVSES</td>
<td>Average social class of students within a school (i.e., school mean for SES).</td>
</tr>
<tr>
<td>HIMKITY</td>
<td>An effects-coded variable (1 = school enrollment exceeds 40% minority; -1 = otherwise).</td>
</tr>
<tr>
<td>AVBACKORD</td>
<td>Average academic background of students within a school (i.e., school mean for BACKGROUND).</td>
</tr>
</tbody>
</table>

NOTE: AVSES and AVBACKORD were constructed from a larger sample of students than those in this analytic sample.
A Random-Effects ANOVA Model. The analysis began with fitting a one-way random-effects ANOVA model in order to determine the total amount of variability in the outcome (senior-year mathematics achievement) within and between schools. The average school mean, \( Y_{00} \), was estimated as 12.125. The pooled within-school or level-1 variance, \( \sigma^2 \), was 39.927, and the variance among the J school means, \( \tau_{00} \), was 9.335. Using these results and Equation 4.6, we can estimate the proportion of variance between schools (i.e., the intraclass correlation) as 0.189. We note that the estimate of \( \sigma^2 \) from the random-effects ANOVA model represents the total level-1 variance.

As we will see below, some of this variance is explained as predictors are introduced into the level-1 model.

A Random-Coefficient Regression Model. The next step in the analysis involved posing a model to represent the social distribution of achievement in each of the J schools. Specifically, at level 1 (the student model), the mathematics achievement for student \( i \) in school \( j \) (\( Y_{ij} \)) was regressed on minority status (MINORITY), social class (SES), and academic background (BACKGROUND):

\[
Y_{ij} = \beta_{0j} + \beta_{1j}(\text{MINORITY})_{ij} + \beta_{2j}(\text{SES})_{ij} + \beta_{3j}(\text{BACKGROUND})_{ij} + r_{ij} \tag{5.27}
\]

Note that the variance of \( r_{ij}, \sigma^2 \), now represents the residual variance at level 1 that remains unexplained after taking into account students’ minority status, social status, and academic background.

Each school’s distribution of achievement is characterized in terms of four parameters: an intercept and three regression coefficients. The MINORITY, SES, and BACKGROUND variables were all group-mean centered (see Chapter 2). As a result, the four parameters can be interpreted as follows:

- \( \beta_{0j} \) is the mean achievement in school \( j \);
- \( \beta_{1j} \) is the “minority” gap in school \( j \) (i.e., the mean difference between the achievement of white and minority students);
- \( \beta_{2j} \) is the differentiating effect of social class in school \( j \) (i.e., the degree to which SES differences among students relate to senior-year achievement); and
- \( \beta_{3j} \) is the differentiating effect of academic background in school \( j \) (i.e., the degree to which differences in students’ academic BACKGROUND eventuate in senior-year achievement differences).

Each of the differentiating effects, \( \beta_{0j}, \beta_{1j}, \beta_{2j}, \text{ and } \beta_{3j} \), are net of the others. For example, the minority gap in school \( j, \beta_{1j} \), is the adjusted mean achievement difference between white and minority students in school \( j \) after controlling for the effects of individual student’s SES and BACKGROUND.

In terms of this model, an effective and equitable school would be characterized by a high level of mean achievement (i.e., a large positive value for \( \beta_{0j} \)), a small minority gap (i.e., a near-zero value for \( \beta_{1j} \)), and weak differentiating effects for social class and academic background (i.e., small positive values for \( \beta_{2j} \) and \( \beta_{3j} \), respectively).

Each of the four coefficients in Equation 5.27 was specified as random in the level-2 model. Specifically,

\[
\beta_{ej} = \gamma_{0e} + u_{ej} \quad \text{for } q = 0, 1, 2, 3, \tag{5.28}
\]

where \( \gamma_{0e} \) is the mean value for each school effect. Because there are four level-2 random effects, the variances and covariances among them now form a 4 by 4 matrix:

\[
T = \begin{bmatrix}
\text{Var}(u_{0j}) & \text{Cov}(u_{0j}, u_{ij}) & \text{Cov}(u_{0j}, u_{2j}) & \text{Cov}(u_{0j}, u_{3j}) \\
\text{Cov}(u_{0j}, u_{ij}) & \text{Var}(u_{1j}) & \text{Cov}(u_{1j}, u_{2j}) & \text{Cov}(u_{1j}, u_{3j}) \\
\text{Cov}(u_{2j}, u_{0j}) & \text{Cov}(u_{2j}, u_{1j}) & \text{Var}(u_{2j}) & \text{Cov}(u_{2j}, u_{3j}) \\
\text{Cov}(u_{3j}, u_{0j}) & \text{Cov}(u_{3j}, u_{1j}) & \text{Cov}(u_{3j}, u_{2j}) & \text{Var}(u_{3j})
\end{bmatrix}
\]

The random-coefficient regression model specified by Equations 5.27 and 5.28 formally represents the hypothesis that the social distribution of achievement, as defined here, varies across the J schools. As shown below, the diagonal elements of the \( T \) matrix provide empirical evidence for examining this hypothesis.

In general, estimation of a random-coefficient regression model is an important early step in a hierarchical analysis. The results from this model guide the final specification of the level-1 equation and provide a range of useful statistics for subsequent model building at level 2.

Table 5.6 presents the results reported by Lee and Bryk (1989). As in the random-effects ANOVA, the average school achievement was estimated as 12.125. The average minority gap, \( \hat{\gamma}_{10} \), was -2.78 points. This means that in a typical school, minority students were scoring 2.78 points behind white schoolmates with academic and social backgrounds like their own. Similarly, student SES and BACKGROUND (\( \hat{\gamma}_{20} \) and \( \hat{\gamma}_{30} \), respectively) were positively related to achievement. This means that in the average high school, more affluent students and those who entered better prepared had higher math achievement in their senior year. The reported \( t \) ratios are quite large, indicating that each of the level-1 predictors was statistically significant.
TABLE 5.6 Random-Coefficient Regression Model of the Social Distribution of Mathematics Achievement

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>se</th>
<th>t Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>School mean achievement,  ( \gamma_{00} )</td>
<td>12.125</td>
<td>0.252</td>
<td>48.207</td>
</tr>
<tr>
<td>Minority gap,  ( \gamma_{01} )</td>
<td>-2.780</td>
<td>0.242</td>
<td>-11.515</td>
</tr>
<tr>
<td>SES differentiation,  ( \gamma_{02} )</td>
<td>1.135</td>
<td>0.104</td>
<td>10.882</td>
</tr>
<tr>
<td>Academic differentiation, ( \gamma_{03} )</td>
<td>2.582</td>
<td>0.093</td>
<td>27.631</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Variance Component</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean achievement,  ( u_{0j} )</td>
<td>9.325</td>
<td>137</td>
<td>1,770.70</td>
<td>0.000</td>
</tr>
<tr>
<td>Minority gap,  ( u_{1j} )</td>
<td>1.367</td>
<td>137</td>
<td>161.01</td>
<td>0.079</td>
</tr>
<tr>
<td>SES differentiation,  ( u_{2j} )</td>
<td>0.360</td>
<td>137</td>
<td>173.39</td>
<td>0.019</td>
</tr>
<tr>
<td>Academic differentiation,  ( u_{3j} )</td>
<td>0.496</td>
<td>137</td>
<td>219.02</td>
<td>0.000</td>
</tr>
<tr>
<td>Level-1 effect,  ( \epsilon_{ij} )</td>
<td>31.771</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correlation Among School Effects</th>
<th>Mean Achievement</th>
<th>Minority Gap</th>
<th>SES Differentiation</th>
<th>Academic Differentiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minority gap</td>
<td>0.387</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SES differentiation</td>
<td>0.182</td>
<td>-0.109</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic differentiation</td>
<td>0.327</td>
<td>0.085</td>
<td>0.652</td>
<td></td>
</tr>
</tbody>
</table>

Reliability of OLS Regression-Coefficient Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean achievement</td>
<td>0.992</td>
</tr>
<tr>
<td>Minority gap</td>
<td>0.098</td>
</tr>
<tr>
<td>SES differentiation</td>
<td>0.167</td>
</tr>
<tr>
<td>Academic differentiation</td>
<td>0.330</td>
</tr>
</tbody>
</table>

The estimated variances of the random effects at levels 1 and 2 (\( \sigma^2 \) and \( \tau_{0q} \), respectively) are reported in the second panel of Table 5.6. Note that the level-1 variance has been reduced from 39.927 in the random-effects ANOVA model to 31.771 after taking into account students’ minority status, social class, and academic background. The proportion of variance explained by this level-1 model is

\[
\frac{(39.927 - 31.771)}{39.927} = 0.204.
\]

The estimated level-2 variances for the random-coefficient regression model provide empirical evidence about the variability in the social distribution of achievement across schools. The homogeneity of variance tests for these level-2 random effects (see Equation 3.103) can be used to test whether the structure of the social distribution of achievement differs across schools. That is, rejecting the hypotheses that

\[
H_0: \text{Var}(u_{0j}) = \text{Var}(\beta_{0j}) = 0 \quad \text{for} \quad q = 0, 1, 2, 3 \tag{5.29}
\]

implies variation among schools in their social distribution of achievement.

In terms of the univariate \( \chi^2 \) tests, the probability of the estimated variability in the \( \beta_{0j} \) coefficients, under a homogeneity hypothesis, is less than .001 for average achievement and academic differentiation, and less than .02 for the SES differentiation. The \( p \) value associated with the hypothesis of slope homogeneity for the minority gap coefficients is marginal (.079). Because substantial differences between sectors in minority achievement had been previously reported, however, this effect was maintained as random by Lee and Bryk.

We note that these \( \chi^2 \) tests provide only approximate probability values for two reasons. First, they are simple univariate tests that do not take into account the other random effects in the model. Second, they are estimated on the basis of only those schools that have sufficient data to compute a separate OLS regressions. In this particular application, only 138 of the total of 160 schools could be used, because the remaining 22 schools had no variation on minority status, which is why \( df = 137 \) in the second panel of Table 5.6.

When in doubt, the results of these univariate homogeneity tests can be cross-checked through the use of a multivariate likelihood-ratio test (see Equations 3.105 to 3.107), which uses all of the data available. Specifically, the deviance statistic from the full random-coefficient regression model can be compared with the corresponding statistic from a restricted model, say for example, a model with only a random intercept:

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

\[
\beta_{0j} = \gamma_{0j} \quad \text{for} \quad q = 1, 2, 3.
\tag{5.30}
\]

In the Lee and Bryk (1989) data, the deviance statistic for the full random-coefficient regression model was 58,248.4 with 11 df. For the restricted model (which specified all regression slopes as fixed), it was 58,283.6 with 2 df. As a result, the likelihood-ratio test statistic was 35.2 with 9 df \((p < .001)\), which offers confirming evidence that schools do vary in their distributive effects.

Plausible value estimates (see Equation 4.19) from the random-coefficient regression model provide useful descriptive statistics of how much schools really vary in terms of mean achievement, size of the minority gaps, and social- and academic-differentiation effects. Under the normality assumption, we would expect the effects for 95% of the schools to fall within the range

\[
\hat{\gamma}_{0j} \pm 1.96(\hat{\tau}_{0j})^{1/2}.
\tag{5.31}
\]
Thus, in this High School and Beyond Survey data, school means ($\beta_{0j}$) would be expected in the range of (6.140, 18.110). Minority gaps ($\beta_{1j}$) of (-5.072, -0.488) are quite plausible, as are social- and academic-differentiation effects of (-0.041, 2.311) and (1.202, 3.962), respectively. Clearly, these results suggest considerable variation among schools on each effect. Interestingly, we could expect to find some schools where minority performance approximates white achievement and where social-class differences are inconsequential because values near zero are plausible for both $\beta_{1j}$ and $\beta_{2j}$. However, all schools appear to engage in some degree of academic differentiation in that values of zero for $\beta_{3j}$ do not appear plausible. We note that the plausible values are for the true school parameters, $\beta_{4j}$, and not the separate OLS estimates of these parameters, $\hat{\beta}_{4j}$. The OLS estimates, especially of the regression slopes, would be far more variable because of the unreliability of sample estimates of these individual school parameters.

Another useful set of descriptive statistics that can be computed from the level-2 variance-covariance components are the correlations among the school effects. For any two random effects $u_{4j}$ and $u_{4j'}$ (or, equivalently, in the random-coefficient regression model $\beta_{4j}$ and $\beta_{4j'}$), respectively,

$$
\rho(u_{4j}, u_{4j'}) = \hat{\tau}_{44}/(\hat{\tau}_{4q} \hat{\tau}_{q4q'})^{1/2}.
$$

These results are reported in the third panel of Table 5.6. Schools displaying high levels of achievement tended to have small minority gaps ($\rho_{01} = 0.397$) but were somewhat more differentiating with regard to social class ($\rho_{02} = 0.182$) and academic background ($\rho_{03} = 0.327$) than schools with lower achievement levels. Interestingly, the social- and academic-differentiation effects were correlated 0.652, suggesting that these two school effects may share some common causes.

In general, it is important to inspect the correlations estimated from the random-coefficient regression model. Although social and academic differentiation were moderately strongly correlated in this application, there was still sufficient independent variation to treat each of them as separate school effects. In applications discussed later in this book, correlations of .90 and higher were found. In such cases, the two random effects are carrying essentially the same variation across the level-2 units. A reduction of the model to specify one of these level-1 effects as fixed or nonrandomly varying would be warranted. Theory and research purposes should dictate which of the two is more important to treat as a random effect.

Table 5.6 also reports the reliabilities for each of the level-2 random effects. In the random-coefficient regression model, these are equivalent to the reliabilities of the OLS estimates, $\hat{\beta}_{4j}$, as measures of the true parameters.

$\beta_{4j}$ These reliabilities were computed by substituting the estimated values for the level-1 and level-2 variance components into Equation 3.59. We note that these statistics, like the $\chi^2$ homogeneity statistics, use the separate OLS estimates for each level-2 unit. Thus, they are based on 138 schools in this application.

The reliability estimates from the random-coefficient regression model are helpful in that they provide additional guidance on appropriate specification of the level-1 coefficients (i.e., as fixed, random, or nonrandomly varying). Because the metric of $\tau_{4q}$ depends on the metric of the corresponding $X_q$ and $Y_q$, interpreting the absolute values of $\tau_{4q}$ is somewhat easier. The reliability provides an alternative indicator of amount of signal present in these data. That is, it tells us how much of the observed variation in the $\hat{\beta}_{4j}$ is potentially explainable. Past experiences working with these methods suggest that whenever the reliability of a random level-1 coefficient drops below 0.05, that coefficient is a candidate for treatment either as fixed or nonrandomly varying.

These statistics also offer insight into the power of a particular data set to detect hypothesized structural effects. We have considerable power in the High School and Beyond data for examining hypotheses about effects of school characteristics on school mean achievement since the intercept estimates are highly reliable. In contrast, the data set is only marginally useful for studying how school characteristics influence the relative achievement of majority and minority children. As noted above, 22 schools have no information on this effect and some of the others have only limited information. This suggests caution in inferring that “school characteristics don’t seem to matter” in terms of influencing the relative achievement levels of majority and minority group children. The reliability coefficients tell us that these data provide little evidence for making such assertions. In short, they caution us against overzealous interpretation of a null hypothesis affirmed.

An Intercepts- and Slopes-as-Outcomes Model: The Effects of Sector and Context. The results from the random-coefficient regression model indicated that each of the level-1 predictors had, on average, a significant relationship with math achievement. (This judgment is based on the fixed-effect estimates, their standard errors, and $t$ ratios.) Thus, each of these predictors should remain at least as a fixed effect in the student-level model. Further, the statistical evidence provided by the $\tau_{4q}$ point estimates, the $\chi^2$ homogeneity tests, the likelihood-ratio test, and the reliability statistics indicated that there was sufficient variability among schools in each of the level-1 regression coefficients to treat these coefficients, at least initially, as random.

Lee and Bryk (1989) next sought to develop explanatory models to illuminate how differences among schools in their organizational characteristics might influence the social distribution of achievement within schools.
One model hypothesized differential effects of sector and composition. The investigators noted that the student composition in both Catholic and public schools varied considerably and that these contextual differences might affect outcomes, even after adjusting for the individual student characteristics already included in the level-1 model. (This idea of compositional effects in organizational research is discussed more fully later.) Thus, they modeled the joint effects of SECTOR and context (as measured by AVSES, HIMNRTY, AVBACKGRD from Table 5.5) on mean achievement, minority gap, social differentiation, and academic differentiation. They also hypothesized that these context effects might be different in the two sectors. Therefore, they included the interactions between SECTOR and each of the context measures as predictors in the level-2 models. The investigators allowed the level-1 model to remain as in Equation 5.27. They posed the following level-2 model:

$$
\begin{align*}
\beta_{0j} &= \gamma_{00} + \gamma_{01}(AVSES)_{ij} + \gamma_{02}(HIMNRTY)_{ij} + \gamma_{03}(AVBACKGRD)_{ij} \\
&\quad + \gamma_{04}(SECTOR)_{ij} + \gamma_{05}(SECTOR \times AVSES)_{ij} \\
&\quad + \gamma_{06}(SECTOR \times HIMNRTY)_{ij} \\
&\quad + \gamma_{07}(SECTOR \times AVBACKGRD)_{ij} + u_{0j}, \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}(HIMNRTY)_{ij} + \gamma_{12}(SECTOR)_{ij} \\
&\quad + \gamma_{13}(SECTOR \times HIMNRTY)_{ij} + u_{1j}, \\
\beta_{2j} &= \gamma_{20} + \gamma_{21}(AVSES)_{ij} + \gamma_{22}(SECTOR)_{ij} \\
&\quad + \gamma_{23}(SECTOR \times AVSES)_{ij} + u_{2j}, \\
\beta_{3j} &= \gamma_{30} + \gamma_{31}(AVBACKGRD)_{ij} + \gamma_{32}(SECTOR)_{ij} \\
&\quad + \gamma_{33}(SECTOR \times AVBACKGRD)_{ij} + u_{3j},
\end{align*}
$$

[5.33]

In their first analysis using this model, several of the estimated coefficients were trivially small (\(\gamma_{06}, \gamma_{07}, \gamma_{11}, \gamma_{12}, \gamma_{31}, \gamma_{33}\)). Each of the corresponding level-2 predictors were deleted and a reduced model estimated. The results for this are presented in Table 5.7 and discussed below.

**School Mean Achievement.** The average academic background of students (AVBACKGRD) was positively related to school mean achievement (\(\bar{y}_{00} = 1.301, t = 2.514\)). Mean achievement was lower in schools with high minority concentrations (\(\bar{y}_{02} = -1.488, t = -2.699\)). The effect of average social class (AVSES) on school mean achievement varied across the two sectors. That is, a significant interaction effect was detected (\(\bar{y}_{03} = -1.572, t = -3.642\)). In the Catholic sector, the relationship of AVSES with school mean achievement was 2.534 [i.e., \(\bar{y}_{00} + (1)\bar{y}_{03} = 4.106 - 1.572\)]. In the public sector, the relationship was much stronger, at 5.678 [i.e., \(\bar{y}_{00} + (-1)\bar{y}_{03} = 4.106 + 1.572\)].

| TABLE 5.7 Estimated Effects of Sector and Control on the Social Distribution of Achievement |
|----------------------------------------|----------------|------------------|----------------|
| Fixed Effect                          | Coefficient  | se            | t Ratio        |
| School mean achievement                |               |                |                |
| BASE, \(\gamma_{00}\)                  | 13.678        | 0.186          | 73.393         |
| AVSES, \(\gamma_{01}\)                 | 4.106         | 0.493          | 8.327          |
| HIMNRTY, \(\gamma_{02}\)               | -1.488        | 0.551          | -2.699         |
| AVBACKGRD, \(\gamma_{03}\)             | 1.301         | 0.517          | 2.514          |
| SECTOR, \(\gamma_{04}\)                | 0.716         | 0.154          | 4.700          |
| SECTOR \times AVSES, \(\gamma_{05}\)   | -1.572        | 0.432          | -3.642         |
| Minority gap                           |               |                |                |
| BASE, \(\gamma_{10}\)                  | -2.894        | 0.256          | -11.300        |
| SECTOR, \(\gamma_{11}\)                | 0.721         | 0.256          | 2.816          |
| Social class differentiation           |               |                |                |
| BASE, \(\gamma_{20}\)                  | 1.381         | 0.141          | 9.819          |
| AVSES, \(\gamma_{21}\)                 | 0.131         | 0.325          | 0.402          |
| SECTOR, \(\gamma_{22}\)                | -0.362        | 0.141          | -2.571         |
| SECTOR \times AVSES, \(\gamma_{23}\)   | -0.869        | 0.325          | -2.671         |
| Academic differentiation               |               |                |                |
| BASE, \(\gamma_{30}\)                  | 2.482         | 0.093          | 26.650         |
| SECTOR, \(\gamma_{31}\)                | 0.072         | 0.093          | 0.778          |

**Random Effect**

<table>
<thead>
<tr>
<th>Variance Component</th>
<th>df</th>
<th>(\chi^2)</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean achievement</td>
<td>2.681</td>
<td>132</td>
<td>631.19</td>
</tr>
<tr>
<td>Minority gap</td>
<td>0.624</td>
<td>136</td>
<td>151.04</td>
</tr>
<tr>
<td>SES differentiation</td>
<td>0.218</td>
<td>134</td>
<td>159.94</td>
</tr>
<tr>
<td>Academic differentiation</td>
<td>0.475</td>
<td>136</td>
<td>221.70</td>
</tr>
<tr>
<td>Level-1 effect, (\tau_0)</td>
<td>31.778</td>
<td>31</td>
<td>31.778</td>
</tr>
</tbody>
</table>

The presence of such an interaction effect means that the magnitude of the sector effect depends on the social class of the schools compared. In general, the sector effect on mean achievement was

\[
\text{Catholic prediction} - \text{public prediction} = (1)\bar{y}_{04} + (1)(AVSES)(\bar{y}_{05}) - [(1)\bar{y}_{04} + (1)(AVSES)(\bar{y}_{05})] = 2[\bar{y}_{04} + (AVSES)(\bar{y}_{05})].
\]

For schools of average social class (AVSES = 0), the sector effect was 1.432 points, or \(2[\bar{y}_{04} + (0)\bar{y}_{05}]\). The Catholic advantage was greater for low social-class schools. For example, if AVSES = -1.0, the sector effect was 4.576 points, or \(2[\bar{y}_{04} + (-1)\bar{y}_{05}]\). For affluent schools, however (AVSES > 1), average mathematic achievement was actually higher in public schools.
Minority Gap. The minority gap was also different in the two sectors ($\hat{y}_{12} = 0.721, t = 2.816$). In the average Catholic school, minority students scored 2.173 points behind their white classmates ($\hat{y}_{10} + (1)\hat{y}_{12}$). (This is a net effect after controlling for students’ SES and BACKGROUND.) In the average public school, the minority gap was 3.615 points ($\hat{y}_{10} + (-1)\hat{y}_{12}$).

Social Class Differentiation. The differentiating effect of social class within a school depended jointly on the AVSES and SECTOR. The estimated interaction effect, $\hat{y}_{23}$, was larger in magnitude than either of the main effects, $\hat{y}_{21}$ and $\hat{y}_{22}$. In the public sector, high social-class schools were more differentiating with regard to student social class than were low social-class schools. In the Catholic sector, the opposite was true: High social-class schools were less socially differentiating than were low social-class schools. This can be seen by computing the effect of AVSES on social-class differentiation separately for the Catholic and public schools based on $\hat{y}_{21} + (\text{SECTOR})\hat{y}_{23}$. For public schools, the effect of AVSES is $0.131 + (-1)(-.869) = 1.000$. For Catholic schools, the effect of AVSES is $0.131 + (1)(-.869) = -.738$.

Academic Differentiation. With regard to academic differentiations, there was no evidence of context, sector, or sector-by-context effects.

Auxiliary Statistics. The bottom panel of Table 5.6 reports the estimated-variance components at level 1 and level 2 for the sector-context effects model. The level-1 variance estimate, $\hat{\sigma}^2$, was virtually identical to that reported for the random-coefficient regression model, an expected result because the level-1 models are the same. In general, some slight variation in the estimation of $\hat{\sigma}^2$ may occur, because all fixed and random effects are estimated jointly and each parameter estimate depends on all of the others.

At level 2, each $\hat{\tau}_{qq}$ estimate is now a conditional or residual variance. That is, $\sigma_{ij}$ is a residual school effect unexplained by the level-2 predictors included in the model. In contrast, each $\hat{\tau}_{qq}$ associated with the random-coefficient model was an unconditional variance. Comparison of these conditional variances (Table 5.7) with the unconditional variances (Table 5.6) indicates a substantial reduction in variation once sector and context are taken into account. The proportion variance explained by the sector-context model at level 2, using Equation 4.24, was

\[
\text{Proportion variance explained in } \beta_3 = \frac{\hat{\tau}_{qq}(\text{unconditional}) - \hat{\tau}_{qq}(\text{conditional})}{\hat{\tau}_{qq}(\text{unconditional})}.
\]

These statistics are reported in Table 5.8. Substantial proportions of the variance in average achievement, minority gap, and social differentiation have been explained. Variation among schools in academic differentiation remains virtually unexplained. Because the reliability of the academic-differentiation effects was relatively high (0.330 in Table 5.6), we can be reasonably confident that the substantial differences observed in academic differentiation were probably not related to sector or context, but rather to other factors. In fact, subsequent analysis reported by Lee and Bryk (1989) explained a substantial variation in $\beta_3$, as a function of differences in the academic organization and normative environments of schools.

Returning to Table 5.7, we note that $\chi^2$ statistics for both the minority gap and SES differentiation effect were consistent with the hypothesis that the residual variation in these two school effects is zero. These results, of course, do not mean that this null hypothesis is actually true. Because the researchers had theoretical reasons to investigate whether these distributive effects also varied as a function of school organization and normative environments, they proceeded to estimate additional models with an expanded set of level-2 predictors for each of the four random school effects. Many hypothesized organizational relations were detected. The point is that the homogeneity tests for intercepts and slopes are only a guide and should not substitute for informed judgement.

Applications with Both Random and Fixed Level-1 Slopes

In the interest of clarity of exposition, we have organized this chapter around two distinct classes of applications. In Case 1, only the intercept parameter varied across organizations, with the effects of level-1 predictors, if included, treated as fixed coefficients. In Case 2 just discussed, all level-1 coefficients were treated as random. In fact, many applications are well suited for a model with both random and fixed level-1 coefficients. For example,
suppose we had one randomly varying level-1 slope, $\beta_{ij}$, and a series of other level-1 predictors that we sought to introduce as covariates:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{1,j}) + \sum_{q=2}^{Q} \beta_{qj}(\bar{X}_{q,j} - \bar{X}_{q}) + r_{ij}.$$  \[5.34\]

At level 2,

$$\beta_{0j} = \gamma_{00} + \sum_{r=1}^{s} \gamma_{0r}W_{ij} + u_{0j} \quad \text{for } q = 0, 1 \quad [5.35a]$$

and

$$\beta_{qj} = \gamma_{q0} \quad \text{for } q = 2, \ldots, Q. \quad [5.35b]$$

where $u_{0j}$ and $u_{qj}$ are assumed bivariate normal with means of 0, variances $\tau_{00}$ and $\tau_{11}$, respectively, and covariances $\tau_{01}$.

In this model, two random effects, an intercept and one slope, are hypothesized for each organization. The intercept, $\beta_{0j}$, is an adjusted mean taking into account differences among the individuals in these organizations with regard to $X_2, \ldots, X_Q$. Similarly, the slope $\beta_{1j}$ is net of any fixed effects associated with $X_2, \ldots, X_Q$. At level 2, $\beta_{0j}$ and $\beta_{1j}$ are hypothesized to vary as a function of measured organizational features, $W_{ij}$.

In general, any combination of random, nonrandomly varying, and fixed coefficients can be employed in the level-1 model. Theoretical considerations are primary in deciding whether a level-1 coefficient should be conceived as random.

**Special Topics**

The remainder of the chapter consists of a series of "special topics" concerning the design and use of hierarchical models in organizational applications. Much of the material presented here is new to the second edition. We demonstrate how the basic hierarchical model can be generalized to represent and model heterogeneous variances at level 1. Following this, we detail how the choice of centering for level-2 variables affects the estimation of random level-1 coefficients, $\beta_{ij}$, fixed effects, $\gamma$, and the variance-covariance components in $T$. We then discuss some complications that can arise in interpreting proportion in variance-explained statistics in more complex organizational effects models. Following this, we describe how to use the empirical Bayes estimates of level-1 coefficients as performance indicators for specific organizational units and discuss validity concerns that may arise in such applications. The chapter concludes with an introduction to power considerations in designing new data collection for two-level organizational effects studies.

**Applications with Heterogeneous Level-1 Variance**

All of the applications considered so far in this chapter have assumed a homogenous residual error at level 1, that is, $\text{Var}(e_{ij}) = \sigma^2$. There is some statistical evidence which suggests that the estimation of the fixed effects, $\gamma$, and their standard errors will be robust to violations of this assumption (Kasin & Raudenbush, 1998). Nonetheless, situations can arise where notable heterogeneity occurs at level 1, which may be substantively significant. In such cases, the analyst may wish to model this heterogeneity as a function of measured variables. These predictors may be defined at either level 1 or level 2.

For example, the High School and Beyond (HS&B) data on students' mathematics achievement, used in Chapter 4, displays heterogeneous residual variance at level 1. (Procedures for detecting heterogeneity of level-1 variance are described in Chapter 9 and illustrated with the HS&B data.) We might hypothesize that the residual variance at level 1 in these data is different for public and Catholic schools. Alternatively, the residual variance might depend on some level-1 characteristic such as gender. (There is some research evidence, for example, that the academic achievement is more variable among boys than girls [Hedges & Nowell, 1995].)

Formally, since the level-1 variance, $\sigma^2$, is constrained to be positive, it is more sensible to model the $\ln(\sigma^2)$ as a linear function of some measured level-1 or level-2 variables, $X$ or $W$, rather than to model $\sigma^2$ itself. This assures that estimates are consistent with only positive values of $\sigma^2$.

Therefore, we now add an additional structural specification to the basic hierarchical linear model:

$$\ln(\sigma^2) = \alpha_0 + \sum \alpha_j C_j. \quad [5.36]$$

where $C_j$ may be either a level-1 or a level-2 predictor, $X$ or $W$, respectively. Since the $\alpha_j$ coefficients in Equation 5.36 will be estimated through maximum likelihood, the coefficients relative to their standard errors form $z$ statistics under large-sample theory. This test statistic can be used, as illustrated below, for purposes of testing hypotheses about the sources of level-1 heterogeneity.

**Example: Modeling Sector Effects on the Level-1 Residual Variance in Mathematics Achievement**

To illustrate this extension of hierarchical linear models using the HS&B data, we consider whether the level-1 residual variance might be different in Catholic and public schools. We pose a simple model with student social class, $X_{ij}$, group mean centered at level 1, and SECTOR as a level-2 predictor.
for both school mean achievement, $\beta_{0j}$, and the SES differentiation effect, $\beta_{1j}$. That is,

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{.j}) + r_{ij}$$

and

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{SECTOR}) + u_{0j},$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{SECTOR}) + u_{1j},$$

where now $r_{ij}$ is assumed to be distributed $N(0, \sigma_r^2)$. In this particular example, we specify that

$$\ln(\sigma_r^2) = \alpha_0 + \alpha_1(\text{SECTOR}).$$

Table 5.9 presents the results for this model estimated both with and without the heterogeneous variance specification at level 1. There is clear evidence that the residual level-1 variability, after controlling for student social class, is different within public and Catholic high schools ($\alpha_1 = -1.82$, with $z = 5.409$). By exponentiating the results from the model for $\ln(\sigma^2)$, we can compute separate level-1 residual variance estimates for Catholic and public schools. The homogeneous model produced an overall level-1 residual variance estimate of 36.705. We see that the residual level-1 variance in Catholic schools is 33.31 while in public schools it is considerably higher, 39.96.

In general, the heterogeneous model appears to fit these data better than the simple homogeneous level-1 specification. In addition to the highly significant $z$ test statistic for $\alpha_1$, this inference is also confirmed in the results from the likelihood-ratio test that compares the deviances from the two models, $\chi^2 = 29.139$, with 1 df, and $p < .000$. Even so, we note that the parameter estimates from the two models for $\gamma$, the $se(\gamma)$ and $T$, remain quite similar.

### Data-Analytic Advice About the Presence of Heterogeneity at Level 1

The potential to model level-1 residual variances represents a useful extension to the basic model for some substantive applications. As discussed in Chapter 9 (see also Raudenbush & Bryk, 1987), the presence of heterogeneity of variance at level 1 can be viewed as an omnibus indicator of level-1 model misspecification. That is, heterogeneity will result at level 1 when an important level-1 predictor has been omitted from the model or the effect of a level-1 predictor has been erroneously specified as fixed, when it should

**TABLE 5.9 Comparison of Homogeneous and Heterogeneous Level-1 Variance Models for Mathematics Achievement from High School and Beyond**

<table>
<thead>
<tr>
<th></th>
<th>Results for Homogeneous Variance Model, $\text{Var}(r_{ij}) = \sigma^2$</th>
<th></th>
<th>Results for Heterogeneous Variance Model, $\ln[\text{Var}(r_{ij})] = \alpha_0 + \alpha_1(\text{SECTOR})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fixed Effect</td>
<td>Coefficient</td>
<td>$se$</td>
</tr>
<tr>
<td>School mean achievement</td>
<td>BASE, $\gamma_{00}$</td>
<td>11.394</td>
<td>0.291</td>
</tr>
<tr>
<td></td>
<td>SECTOR, $\gamma_{10}$</td>
<td>2.807</td>
<td>0.436</td>
</tr>
<tr>
<td>Social class differentiation</td>
<td>BASE, $\gamma_{01}$</td>
<td>2.803</td>
<td>0.154</td>
</tr>
<tr>
<td></td>
<td>SECTOR, $\gamma_{11}$</td>
<td>-1.341</td>
<td>0.232</td>
</tr>
<tr>
<td>Random Effect</td>
<td>Standard Deviation</td>
<td>$\sigma_r$</td>
<td>Variance Component</td>
</tr>
<tr>
<td>Mean achievement, $u_{0j}$</td>
<td>2.577</td>
<td>6.641</td>
<td>158</td>
</tr>
<tr>
<td>SES differentiation, $u_{1j}$</td>
<td>0.489</td>
<td>0.239</td>
<td>158</td>
</tr>
</tbody>
</table>

Number of estimated parameters = 8

<table>
<thead>
<tr>
<th></th>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>$se$</th>
<th>$t$ Ratio</th>
<th>df</th>
<th>$p$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>School mean achievement</td>
<td>BASE, $\gamma_{00}$</td>
<td>11.394</td>
<td>0.291</td>
<td>39.166</td>
<td>158</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>SECTOR, $\gamma_{10}$</td>
<td>2.807</td>
<td>0.436</td>
<td>6.434</td>
<td>158</td>
<td>0.000</td>
</tr>
<tr>
<td>Social class differentiation</td>
<td>BASE, $\gamma_{01}$</td>
<td>2.803</td>
<td>0.154</td>
<td>18.215</td>
<td>158</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>SECTOR, $\gamma_{11}$</td>
<td>-1.341</td>
<td>0.232</td>
<td>-5.777</td>
<td>158</td>
<td>0.000</td>
</tr>
<tr>
<td>Random Effect</td>
<td>Standard Deviation</td>
<td>$\sigma_r$</td>
<td>Variance Component</td>
<td>df</td>
<td>$\chi^2$</td>
<td>$p$ Value</td>
</tr>
<tr>
<td>Mean achievement, $u_{0j}$</td>
<td>2.574</td>
<td>6.626</td>
<td>158</td>
<td>1,392.58</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>SES differentiation, $u_{1j}$</td>
<td>0.499</td>
<td>0.249</td>
<td>158</td>
<td>173.52</td>
<td>0.189</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>$se$</th>
<th>$z$ Ratio</th>
<th>$p$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept, $\alpha_0$</td>
<td>3.688</td>
<td>0.024</td>
<td>154.533</td>
<td>0.000</td>
</tr>
<tr>
<td>SECTOR, $\alpha_1$</td>
<td>-0.182</td>
<td>0.033</td>
<td>-5.409</td>
<td>0.000</td>
</tr>
</tbody>
</table>

For public schools, $\hat{\sigma}^2 = \exp[3.688 - 0.182(0)] = 39.96$
For Catholic schools, $\hat{\sigma}^2 = \exp[3.688 - 0.182(1)] = 33.31$

### Summary of Model Fit

<table>
<thead>
<tr>
<th></th>
<th>Number of Parameters</th>
<th>Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Homogeneous level-1 variance</td>
<td>8</td>
<td>4.6632.04</td>
</tr>
<tr>
<td>2. Heterogeneous level-1 variance</td>
<td>9</td>
<td>4.6602.90</td>
</tr>
</tbody>
</table>

$$\chi^2$$ | df | $p$ Value |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 versus Model 2</td>
<td>29.139</td>
<td>1</td>
</tr>
</tbody>
</table>
be treated as either random or nonrandomly varying. Thus, the sector differences for \( \sigma^2 \) found in the application above suggest that we pursue further the specification of the level-1 model. What could cause these differences in residual variances between Catholic and public schools? Two immediate explanations arise: (1) The student intake into Catholic and public high schools might differ on other factors, in addition to social class, and these student-level background controls need to be added to the model; and/or (2) students might have different academic experiences within Catholic and public high schools that contribute to a different final distribution of achievement in these two sectors. For example, students pursue more differentiated course taking patterns in public schools, which could result in the greater heterogeneity found there (Bryk, Lee, & Holland, 1993). This explanation suggests adding a student-level measure of course taking to the model. In either instance, as we fill out the level-1 model, we would expect the residual heterogeneity to decline.

In short, the presence of heterogeneity at level 1 suggests a need for further modeling efforts at this level. The possibility exists that the model is misspecified in ways that might bias the estimates of \( \gamma \) and \( T \). Moreover, while a heterogeneous specification for \( \sigma^2 \) at level 1 can help provide clues about the sources of this heterogeneity, it does not per se protect against model misspecification bias. Only a more elaborated level-1 model can help in this regard.

**Centering Level-1 Predictors in Organizational Effects Applications**

As discussed in Chapter 2 (see "Location of Xs"), choice of location for level-1 predictors affects the definition of the level-1 intercept in two-level models. In some applications, centering around a constant such as the grand mean of \( X \) is advisable, while in other settings, centering around a level-2 mean (i.e., "group-mean centering") will be preferable. We now consider how these choices affect inferences for five purposes:

- estimating fixed level-1 coefficients;
- disentangling person-level and compositional effects;
- estimating level-2 effects while adjusting for level-1 covariates;
- estimating the variances of level-1 coefficients; and
- estimating random level-1 coefficients.

**Estimating Fixed Level-1 Coefficients**

In addition to estimating how organizational factors influence person-level outcomes, multilevel data are often also used to estimate person-level effects. For example, in the High School and Beyond data, a primary analytic concern was the relationship between student social class and math achievement. The fact that students were nested within schools represented a nuisance consideration in an effort to obtain an appropriate estimate of this level-1 relationship. Burstein (1980) provided an extensive review of the modeling issues that arise in such applications. We offer below a brief account of these concerns and show how they can be resolved through the formulation of hierarchical models that explicitly represent the nesting structure. These methods are illustrated using the High School and Beyond data. We assume for now that the level-1 relationship of interest is fixed across level-2 units.

We begin by considering the most commonly employed technique for analyzing multilevel data—an OLS regression analysis at level 1, which simply ignores the nesting of persons within groups. The first column of Table 5.10 presents the model for this analysis, and the results from a regression of math achievement on student SES. The estimated regression coefficient for SES was 3.184, with a standard error of 0.097. For reasons that will become clear below, we refer to the regression coefficient in this model as \( \beta \).

For comparison, we present in the second column of Table 5.10 the corresponding model for a level-2 or between-group analysis. When person-level data are not available, this regression coefficient, referred to as \( \beta_n \), has often been used as an estimator of the person-level relationship. (Robinson [1950] Burstein [1980] and Aitkin and Longford [1986] discuss the conditions under which such use is appropriate.) In this particular application, however, the estimated \( \beta_n \), of 5.909 is almost twice as large as \( \beta \). Clearly, these two analyses provide very different answers about the magnitude of the relationship between individual social class and math achievement. Notice also that the standard error is considerably larger in this case (0.371 versus 0.097), primarily reflecting the fact that the degrees of freedom in the level-2 analysis are 158, compared with 7,183 in the level-1 analysis. Even when it is appropriate to use \( \beta_n \) as an estimator of the person-level relationship (which does not appear to be the case here), the estimate will generally be less precise than an estimate based on a level-1 analysis. Typically, there is just less information available in the unit means than in the full individual data.

It is frequently argued (see, e.g., Firebaugh, 1978) that the person-level coefficient really of interest is the pooled-within-organization relationship between math achievement and student SES. That is, we want to estimate the level-1 relationship net of any group-membership effects. This coefficient
is typically referred to as $\beta_w$ and can be obtained with OLS methods by estimating the equation

$$Y_{ij} - \overline{Y}_{..j} = \beta_w (X_{ij} - \overline{X}_{..j}) + r_{ij}. \tag{5.37}$$

Column 3 of Table 5.10 presents the equivalent hierarchical linear model. For the High School and Beyond data, $\beta_w$ was 2.191 with a standard error of 0.109. Although the estimated standard error for $\beta_w$ was quite similar to the standard error for $\hat{\beta}_1$ (which will typically be the case), $\hat{\beta}_1$ was actually partway between $\beta_w$ and $\hat{\beta}_0$. It can readily be shown that $\hat{\beta}_1$ is formally a weighted combination of $\beta_w$ and $\hat{\beta}_0$:

$$\hat{\beta}_1 = \eta^2 \hat{\beta}_0 + (1 - \eta^2) \beta_w, \tag{5.38}$$

where $\eta^2$ is the ratio of the between-schools sum of squares on SES to the total sum of squares on SES.

The relationships among $\beta_w$, $\beta_0$, and $\hat{\beta}_1$ are depicted in Figure 5.1. The figure shows a hypothetical data set with three schools: a school with low mean SES, a school with medium SES, and a school with high mean SES. Within each school is a regression line with slope $\beta_w$ that describes the association between student SES and $Y$ within that school. These within-school slopes are assumed equal. There is also a regression line (the bold line) that describes the association between mean SES and mean $Y$. This is the line one would estimate using only three data points, that is, the mean

![Figure 5.1](image-url)
SES and mean $Y$ for each school. This line has slope $\beta_b$. A final dashed line describes the association between student SES and $Y$, ignoring the clustering of students within schools. The slope $\beta_c$ of this line is neither as flat as $\beta_w$ nor as steep as $\beta_b$.

Thus, ignoring the nested structure of the data can lead to misleading results when person-level effect estimates are desired. As Cronbach (1976) noted, $\hat{\beta}_b$ is generally an uninterpretable blend of $\hat{\beta}_w$ and $\hat{\beta}_c$. In most research applications, the person-level effect estimate of interest is $\hat{\beta}_w$ and not $\hat{\beta}_c$.

In estimating $\hat{\beta}_w$ based on a hierarchical analysis, the group-mean centering of $X_j$ plays a critical role. By contrast, if the data are grand-mean centered, as in the fourth column of Table 5.10, the resulting estimator is a mix of $\hat{\beta}_w$ and $\hat{\beta}_b$. The weights, $W_1$ and $W_2$, are quite complex for the general case. Note that the coefficient estimate for the High School and Beyond data with grand-mean centering was 2.391.

This result derives from the fact that the model with grand-mean centering actually involves both $X_{ij}$ and $\bar{X}_{.,j}$, but the analysis is constrained to estimate only one parameter rather than separate estimates for $\hat{\beta}_w$ and $\hat{\beta}_b$ (a hierarchical model for the joint estimator of $\hat{\beta}_w$ and $\hat{\beta}_b$ is presented in the next section).

It is important to note that if in a given application $\hat{\beta}_b$ and $\hat{\beta}_w$ were in fact identical then the estimated model with grand-mean centering would be most efficient. Under the hypothesis that $\beta_w = \beta_w = \beta$, $\hat{\beta}_b$ and $\hat{\beta}_w$ are independent, unbiased estimators of $\beta$. In a balanced design, the sampling variances of each of these OLS estimates, respectively, is

$$\text{Var}(\hat{\beta}_b) = \Delta / \sum_j (\bar{X}_{.,j} - \bar{X}_{.,j})^2,$$

with

$$\Delta = \tau_{9b} + \sigma^2 / n,$$

and

$$\text{Var}(\hat{\beta}_w) = \sigma^2 / \sum_j \sum_i (X_{ij} - \bar{X}_{.,j})^2.$$

In this specific case, the estimator of $\beta$ provided in column 4 of Table 5.10,

$$\hat{\gamma}_{10} = \hat{\beta} = \frac{W_1\hat{\beta}_b + W_2\hat{\beta}_w}{W_1 + W_2},$$

simplifies in that

$$W_1 = [\text{Var}(\hat{\beta}_b)]^{-1} = \sum_j (\bar{X}_{.,j} - \bar{X}_{.,j})^2 / \Delta$$

and

$$W_2 = [\text{Var}(\hat{\beta}_w)]^{-1} = \sum_j (X_{ij} - \bar{X}_{.,j})^2 / \sigma^2.$$

In this specific case, $\hat{\gamma}_{10}$ is a weighted average of $\hat{\beta}_b$ and $\hat{\beta}_w$ where the weights are the precisions of each estimator. Because both $\hat{\beta}_b$ and $\hat{\beta}_w$ contain information about $\beta$, the hierarchical estimator optimally combines the information to yield a single estimator with greater precision than either of the two component estimators. Formally,

$$[\text{Var}(\hat{\beta}_{\text{hierarchical}})]^{-1} = W_1 + W_2,$$

that is, the precision of the hierarchical estimate is the sum of the precisions of $\hat{\beta}_b$ and $\hat{\beta}_w$. It is of interest that $W_1$ is approximately proportional to $J$ and $W_2$ to $Jn$.

In unbalanced designs, the formulas become more complicated. The same principle still applies, however. Assuming $\beta_w = \beta_b$, the hierarchical estimator with grand-mean centering will be most efficient.

When $\beta_b \neq \beta_w$, as appears true in this application, the hierarchical estimator under grand-mean centering is an inappropriate estimator of the personal-level effect. It too is an uninterpretable blend: neither $\beta_b$ nor $\beta_w$ nor $\beta_b$.

Thus, when an unbiased estimate of $\beta_w$ is desired, group-mean centering will produce it. Two alternative hierarchical models that allow joint estimation of $\beta_w$ and $\beta_b$ are presented below.

**Disentangling Person-Level and Compositional Effects**

Compositional or contextual effects are of enduring interest in organizational sociology (see, e.g., Erbring & Young, 1979; Firebaugh, 1978). Such effects are said to occur when the aggregate of a person-level characteristic, $\bar{X}_{.,j}$, is related to the outcome, $Y_{ij}$, even after controlling for the effect of the individual characteristic, $X_{ij}$. In an OLS level-1 regression analysis, these effects are represented through the inclusion of both $(X_{ij} - \bar{X}_{.,j})$ and $\bar{X}_{.,j}$ as predictors:

$$Y_{ij} = \beta_0 + \beta_1(X_{ij} - \bar{X}_{.,j}) + \beta_2\bar{X}_{.,j} + \epsilon_{ij}.$$  \[5.41\]

The compositional effect is the extent to which the magnitude of the organization-level relationship, $\beta_0$, differs from the person-level effect, $\beta_w$. Formally, the compositional effect is

$$\beta_c = \beta_2 - \beta_1 = \beta_b - \beta_w.$$  \[5.42\]

The compositional effect is graphed in Figure 5.2. We note that a nonzero estimate for $\beta_2$ does not necessarily imply a compositional effect. If $\beta_1$ and $\beta_2$ are equal, no compositional effect is present.

Compositional effects are open to widely varying interpretations. Such effects may occur because of normative effects associated with an organization or because $\bar{X}_{.,j}$ acts as a proxy for other important organizational
Within a hierarchical modeling framework, these effects can be estimated in two different ways. In both cases, the person-level $X_{ij}$ is included in the level-1 model and its aggregate, $\bar{X}_{ij}$, is included in the level-2 model for the intercept. The difference between the two approaches, as displayed in columns 1 and 2 of Table 5.11, is in the choice of centering for $X_{ij}$. When group-mean centering is chosen as in column 1, the relationship between $X_{ij}$ and $Y_{ij}$ is directly decomposed into its within- and between-group components. Specifically, $\gamma_{0i}$ is $\beta_b$ and $\gamma_{00}$ is $\beta_w$. The compositional effect can be derived by simple subtraction. Alternatively, if $X_{ij}$ is centered around the grand mean, as in column 2, the compositional effect is estimated directly and $\beta_b$ is derived by simple addition.

The presence of a context effect is graphically illustrated in Figure 5.2. As before, we present results from three schools, each differing respectively from the other by one unit on mean SES. Also, as before, the within-school relationship, $\beta_w$, is graphed for each school, as is the between-school relationship among the schools, $\beta_b$. $\beta_w$ represents the expected difference in $Y$ between two students in the same school who differ by one unit on $X$. (This is illustrated within School 1.) In contrast, $\beta_b$ is the expected difference between the means of two schools, which differ by one unit on mean SES. (This is illustrated for School 1 versus School 2). The contextual effect, $\beta_c = \beta_b - \beta_w$, is the expected difference in the outcomes between two students who have the same individual SES, but who attend schools differing by one unit in mean SES. As illustrated in Figure 5.2, the contextual effect, $\beta_c$, is the increment to learning that accrues to a student by virtue of being educated in School 2 versus School 1.

In terms of the High School and Beyond data, the student-level effect was 2.191 (the same as in column 3 of Table 5.10), the school-level effect was 5.866, and the difference between these two, the compositional effect, was 3.675. Identical estimates result from the two alternative formulations. Clearly, the social-class composition of the school has a substantial association with math achievement, even larger than the individual student-level association.

We should note that similar point estimates for $\beta_b$, $\beta_w$, and $\beta_c$ can be obtained through use of a level-1 OLS regression based on Equation 5.41. In general, the OLS estimates are unbiased but not as efficient as the hierarchical linear model estimators (see Chapter 3). Also, the OLS standard errors for $\hat{\beta}_b$ and $\hat{\beta}_w$ are negatively biased because Equation 5.41 fails to represent explicitly the random variation among schools captured in the $u_{0ij}$.

---

**TABLE 5.11 Illustration of Person-Level and Compositional (or Contextual) Effects**

<table>
<thead>
<tr>
<th>Group-Mean Centering</th>
<th>Statistical Model</th>
<th>Grand-Mean Centering</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{ij} = \beta_{ij} + \beta_{0j}(X_{ij} - \bar{X}<em>{ij}) + \nu</em>{ij}$</td>
<td>$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}<em>{ij}) + \nu</em>{ij}$</td>
<td>$Y_{ij} = \gamma_{0i} + \gamma_{0j}(X_{ij} - \bar{X}<em>{ij}) + \nu</em>{ij}$</td>
</tr>
<tr>
<td>$\beta_{0j} = \gamma_{0j} = \gamma_{0i}$</td>
<td>$\beta_{1j} = \gamma_{0j}$</td>
<td>$\gamma_{0j} = \beta_w$</td>
</tr>
<tr>
<td>$\beta_{0i} = \gamma_{0i}$</td>
<td>$\gamma_{0j} = \beta_w$</td>
<td>$\beta_b = \gamma_{0i} + \gamma_{0j}$</td>
</tr>
<tr>
<td>$\beta_{cj} = \gamma_{0i} - \gamma_{0j}$</td>
<td>$\beta_{cj} = \beta_c$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates Using High School and Beyond Data</th>
<th>Coefficient</th>
<th>se</th>
<th>Coefficient</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{0b}$</td>
<td>12.648</td>
<td>0.149</td>
<td>12.661</td>
<td>0.149</td>
</tr>
<tr>
<td>$\gamma_{0b}$</td>
<td>5.866</td>
<td>0.362</td>
<td>3.675</td>
<td>0.378</td>
</tr>
<tr>
<td>$\gamma_{0b}$</td>
<td>2.191</td>
<td>0.109</td>
<td>2.191</td>
<td>0.109</td>
</tr>
<tr>
<td>$\beta_b$</td>
<td>3.675</td>
<td>0.378</td>
<td>5.866</td>
<td>0.362</td>
</tr>
</tbody>
</table>

a. Not directly estimated but can be determined from the sampling variance-covariance matrix for the $\gamma$ coefficients.
Estimating Level-2 Effects While Adjusting for Level-1 Covariates

One of the most common applications of HLM in organizational research is simply to estimate the association between a level-2 predictor and the mean of \( Y \), adjusting for one or more level-1 covariates. It is assumed (or established through empirical analysis) that there is no compositional effect. We saw an example of this earlier (see "Evaluating Program Effects on Writing"). In this setting, group-mean centering would be inappropriate. Under the group-mean-centered model, the level-1 intercept is the unadjusted mean of the outcome, denoted here as \( \mu_j \) for emphasis,

\[
Y_{ij} = \mu_j + \beta_{1j}(X_{ij} - \bar{X}_j) + r_{ij}. \tag{5.43}
\]

Note the influence of the level-1 predictor \( X \) disappears when we compute the sample mean for each school:

\[
\bar{Y}_j = \mu_j + \bar{r}_j. \tag{5.44}
\]

A simple level-2 model represents the contribution of a level-2 predictor \( W_j \):

\[
\mu_j = \gamma_{00} + \gamma_{0i}W_j + u_{0j}. \tag{5.45}
\]

In short, under group-mean centering, the effect of \( W_j \) is not adjusted for \( X \) in this analysis.

In contrast, when we grand-mean-center \( X \), the level-1 model becomes

\[
Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_.) + r_{ij}
= \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_j + \bar{X}_j - \bar{X}_.) + r_{ij}
= \beta_{0j} + \beta_{1j}(\bar{X}_j - \bar{X}_.) + \beta_{1j}(X_{ij} - \bar{X}_j) + r_{ij}. \tag{5.46}
\]

A comparison of Equations 5.43 and 5.46 reveals that

\[
\beta_{0j} = \mu_j - \beta_{1j}(\bar{X}_j - \bar{X}_.). \tag{5.47}
\]

Thus, in the grand-mean-centered model, the intercept, \( \beta_{0j} \), is the mean \( \mu_j \) minus an adjustment. Now, viewing this adjusted mean as the outcome, we have, at level 2,

\[
\beta_{0j} = \mu_j - \beta_{1j}(\bar{X}_j - \bar{X}_.) = \gamma_{00} + \gamma_{0i}W_j + u_{0j}, \tag{5.48}
\]

In Equation 5.48, the estimate of the effect of \( W_j \) will be adjusted for differences between organizations in the mean of \( X \), the level-1 explanatory variable.

Estimating the Variances of Level-1 Coefficients

Let us again consider a level-1 model with a single level-1 covariate, \( X \). The key goal of the analysis now is to estimate the variance, \( \tau_{11} \), of the level-1 coefficient, \( \beta_{1j} \). If every organization has the same mean of \( X \), apart from sampling variation and randomly missing data, choice of centering would have no important effect on inference about \( \tau_{11} \). In this case, the analyst can choose the location for \( X \) that gives the most sensible definition of the intercept without concern about the effect of centering on the estimation of the variance of the level-1 slopes.

Estimation of \( \tau_{11} \) becomes considerably more complex, however, when the group mean of \( X \) varies systematically across schools. Such variation will typically arise in organizational research for two reasons. First, persons are often selected or assigned to organizations in ways that segregate those organizations, to some degree, on \( X \). Schools, for example, will be somewhat segregated on the basis of social class, firms will be somewhat segregated on the basis of workers' educational background, neighborhoods will be segregated to some degree on the basis of income and ethnicity. Second, organizations may capitalize on their compositions to create effects. Schools with high average social class may have more positive peer influences on student learning than schools with lower mean social class. Firms with highly educated workers may introduce new technology more quickly than other firms.

In general, if the mean of \( X \) systematically varies across schools, choice of centering (i.e., group-mean centering versus centering around a constant) will make a difference in estimating \( \tau_{11} \). In such situations, we recommend group-mean centering to detect and estimate properly the slope heterogeneity. To develop the rationale for this recommendation, we turn again to the data from High School and Beyond (HS&B).

Recall that in our analysis of math achievement in U.S. high schools, MEAN SES is related to both the intercept and the SES-achievement slope. In this context, let us see how choice of centering affects estimation of the level-2 variances. We begin with an unconditional level-2 model, then turn to a model that includes MEAN SES. In each case, the model uses either group-mean centering, that is,

\[
Y_{ij} = \mu_j + \beta_{1j}(X_{ij} - \bar{X}_j) + r_{ij}, \tag{5.49}
\]

or grand-mean centering, that is,

\[
Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_.) + r_{ij}. \tag{5.50}
\]
TABLE 5.12  Effect of Centering on Estimation of Level-1 Slope Variation

<table>
<thead>
<tr>
<th>Model</th>
<th>Level 1</th>
<th>Level 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( y_{ij} = \beta_{ij} + \beta_{ij}X_{ij} + \tau_{ij} ) ( \tau_{ij} \sim N(0, \sigma^2) )</td>
<td>Unconditional: ( \beta_{ij} = \gamma_{ij} + u_{ij} ) ( \beta_{ij} = \gamma_{ij} + u_{ij} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Conditional: ( \beta_{ij} = \gamma_{ij} + \gamma_{ij} \text{(SECTOR)} + \gamma_{ij} \text{(MEAN SES)} + u_{ij} ) ( \beta_{ij} = \gamma_{ij} + \gamma_{ij} \text{(SECTOR)} + \gamma_{ij} \text{(MEAN SES)} + u_{ij} )</td>
</tr>
</tbody>
</table>

**Unconditional Results**

<table>
<thead>
<tr>
<th>Group-mean centering</th>
<th>Conditional Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\tau} = \begin{bmatrix} 8.68 &amp; 0.65 \ 0.65 &amp; 0.68 \end{bmatrix} )</td>
<td>( \hat{\tau} = \begin{bmatrix} 2.38 &amp; 0.19 \ 0.19 &amp; 0.15 \end{bmatrix} )</td>
</tr>
<tr>
<td>( \hat{\sigma}^2 = 36.70 )</td>
<td>( \hat{\sigma}^2 = 36.70 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grand-mean centering</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\tau} = \begin{bmatrix} 4.83 &amp; -0.15 \ -0.15 &amp; 0.42 \end{bmatrix} )</td>
<td>( \hat{\tau} = \begin{bmatrix} 2.41 &amp; 0.19 \ 0.19 &amp; 0.06 \end{bmatrix} )</td>
</tr>
<tr>
<td>( \hat{\sigma}^2 = 36.83 )</td>
<td>( \hat{\sigma}^2 = 36.74 )</td>
</tr>
</tbody>
</table>

**Unconditional Model.** Based on the unconditional level-2 model, we obtain the variance-covariance estimates given in Table 5.12 (column 1). The variance of the intercept, \( \tau_{00} \), is much larger at 8.68 in the group-mean-centered model than 4.83 in the grand-mean-centered model. This can be explained by the fact that the intercept in the group-mean-centered model is the unadjusted mean, while the intercept in the grand-mean-centered model is the adjusted mean (see previous section). However, we also note that the \( \tau_{11} \) estimates are also quite different, at .68 in the group-mean-centered model and .42 in the grand-mean-centered model. Why should this slope variance differ as a function of centering? Given that MEAN SES and SECTOR are related to both the intercept and the slope, we might speculate that adding these predictors at level 2 will eliminate the discrepancy in the \( \tau_{11} \) estimates as a function of centering.

**Conditional Model.** The results with SECTOR and MEAN SES as predictors at level 2 of both \( \beta_{ij} \) and \( \beta_{ij} \) are given in Table 5.12, column 2. Note that the two estimates of \( \tau_{00} \), the variance of the intercepts, now essentially converge at 2.38 (group-mean-centered model) and 2.41 (grand-mean-centered model). This occurs because both models now incorporate a compositional effect, \( \beta_{ij} \), on the intercept. However, the \( \tau_{11} \) estimates remain quite different, at 0.15 in the group-mean-centered model and 0.06 in the grand-mean-centered model. This occurs even though MEAN SES and SECTOR are incorporated into the model for the slope. Clearly, the choice of centering for X in this application affects our estimates for the \( \tau_{11} = \text{Var}(\beta_{ij}) \).

How Centering Affects Estimation of \( \tau_{11} \). To understand how group- and grand-mean centering affect the estimation of the variance of a random coefficient requires some understanding of how the method of maximum likelihood (ML) works in this setting. Formally, the new ML estimate at each iteration of the EM algorithm (see Equation 3.76) is the sum of two components:

\[
\hat{\tau} = J^{-1} \left[ \sum_{j=1}^{J} \left( \frac{u_{1j}^2}{V_{11j}} \right) \right].
\]

The first component, \( u_{1j}^2 / V_{11j} \), is the square of the empirical Bayes (EB) residual associated with the regression coefficient \( \beta_{ij} \). This EB residual is also known as the expected value of \( u_{1j} \) given the data \( Y \) and current estimates of the model parameters. The second component, \( V_{11j} \), is the posterior variance of \( u_{1j} \) given the data \( Y \) and the current estimates of the model parameters. Similarly, the estimate of \( \tau_{00} \) is based on the squared empirical Bayes residual for the intercept plus the posterior variance of \( u_{0j} \.

These formulas indicate that the estimation of the variances of the slope and intercept depend directly on the empirical Bayes residuals. Group-mean and grand-mean centering, however, create different definitions for these residuals.

This difference can be made vivid by considering Figure 5.3, an idealized version of the HS&B data. The figure shows three types of schools: low-mean-SES schools, average-mean-SES schools, and high-mean-SES schools. Within each type of school, the OLS estimates of the achievement-SES slopes vary somewhat. Under group-mean-centering (not shown in Figure 5.3), the OLS estimates of the intercept will experience some shrinkage. The shrinkage will be large when the estimated intercept has low reliability. However, this shrinkage will not have much influence on the slope, unless the two are highly correlated, which is not the case in the HS&B data. While unreliable OLS slope estimates will certainly experience shrinkage, the direction of the shrinkage will be relatively unaffected by the shrinkage of the intercept.

In contrast, the intercept in a grand-mean-centered model estimates an adjusted mean. A heuristic illustration of the OLS estimates for the adjusted means, \( \beta_{0j} \), for six schools appear in Figure 5.3a. The key point here is that for schools that are either very low or very high on mean SES, the
adjusted means can entail an extrapolation (represented by the dashed lines) well beyond the data actually collected in these schools (represented by the solid lines). That is, the adjusted mean for school \( j \) represents the expected outcome for a child at that school who is at the grand mean on SES. As the figure illustrates, however, Schools 1, 2, 5, and 6 do not enroll any children like this. As a result, their adjusted grand means are less reliably measured than for Schools 3 and 4 (assuming similar sample sizes in all six schools.) Consequently, the empirical Bayes estimates for \( \beta_{ij} \) in Schools 1, 2, 5, and 6 will be shrunk extensively toward the grand mean of \( \bar{Y} \). (Compare the OLS estimated \( \beta_{0j} \) in Figure 5.3a with those in Figure 5.3b). A key consequence of the shrinkage in the adjusted means is a corollary shrinkage in the slopes, \( \beta_{1j} \), as well. Specifically, schools with relatively flat OLS slopes and extreme values on school SES (either high or low) record substantially increased empirical Bayes slope estimates. (Notice the substantial impact on the slope estimate for School 1 for example.) The overall result is a homogenization of the slope variability and underestimation of \( \tau_{11} \).

We now apply these ideas to analyze the effects on estimating \( \tau_{11} \) under grand-mean centering with the HS&B data. We have shown previously that there is considerable variability among schools on mean SES. As a result, the “extrapolation problem” under grand-mean centering identified above should plague both very low and very high SES schools. One additional complication enters in the HS&B data. We also showed previously that the magnitude of the achievement-SES slopes depends on school SES. Specifically, we found that high-SES schools tend to have steeper achievement-SES slopes and low-SES schools have flatter slopes (see Table 4.5). Based on the heuristic example discussed above, we expect the empirical Bayes estimates of the slopes in low-SES schools to experience substantial positive shrinkage under grand-mean centering toward steeper slopes. As for the high-SES schools, because their OLS slopes tend to be steeper here, somewhat less shrinkage is likely to occur for the empirical Bayes slopes in these schools under grand-mean centering. That is, with steeper OLS slopes in high-SES schools, the OLS adjusted means for these schools are already projected down toward \( \bar{Y} \). Although these OLS adjusted means may be just as unreliable for low-SES schools, since they are not as deviant from \( \bar{Y} \), less absolute shrinkage should occur under grand-mean centering. This is exactly what we observe in Figure 5.4.

The vertical axis in Figure 5.4 plots the difference between the EB slope residuals under group- versus grand-mean centering, that is,

\[
Y = u_{ij}^{*\text{group}} - u_{ij}^{*\text{grand}},
\]
while the horizontal axis gives the school mean SES. As predicted, in low SES schools, the empirical Bayes residuals are more negative under group-mean centering than under grand-mean centering. This happens because the grand-mean centering has artificially steepened the slopes for these schools (see Figure 5.3). If the achievement-SES slopes had not been related to school SES, we would have expected the exact same phenomenon to occur in high-SES schools and an overall symmetric curvilinear relationship should have appeared. That is, we would have expected to find more negative residuals in high-SES schools under group-mean centering as compared to grand-mean centering. Negative residuals do occur for high SES schools but not as frequently as with low-SES schools.

In general, the empirical Bayes slope estimates based on grand-mean centering are less credible because they can be significantly perturbed by the extrapolations necessary to compute an adjusted mean for each school. The empirical Bayes intercept estimates can experience significant shrinkage toward the center of the distribution. In the process, the regression slopes are artificially homogenized and the end result is negatively biased estimates of the slope variability. Such perturbation does not occur under group-mean centering. Moreover, as noted previously, incorporating mean SES as a predictor variable for both \( \beta_{0j} \) and \( \beta_{1j} \) does not solve this problem.

We note that this problem with grand-mean centering is not unique to organizational effects studies. It can arise in any hierarchical modeling context where variation exists among units on the mean of a level-1 predictor and where substantial slope variability exists for that predictor. We consider, for example in Chapter 6, how choice of centering can affect estimation of the variance of growth rates in a longitudinal context.

**Estimating Random Level-1 Coefficients**

We saw in the previous section how centering affects the estimation of the variance of level-1 regression coefficients. The distortions that arose in estimating this variance arose from the empirical Bayes estimates of the intercept and slope. The same issues are clearly present, therefore, in the estimation of the unit-specific regression equations. The following conclusions appear justified:

1. If the level-1 sample size is large for a given level-2 unit, OLS and EB estimates of that unit's regression function will converge. This convergence does not depend on the method of centering chosen for \( X \).

2. When the level-1 sample size is small or moderate, EB estimates will tend to be more stable and have smaller mean-squared error of estimation than will the OLS estimates. In this case, there are two possibilities:
   a. If the group mean of \( X \) is invariant across level-2 units, or if its variation is ignorable (representing only sampling fluctuations or randomly missing data), group-mean versus grand-mean centering should produce similar results, though grand-mean centering may add a modicum of precision.
   b. If the group mean of \( X \) varies substantially, group-mean centering is likely to produce more robust estimates of unit-specific regression equations than is grand-mean centering.

**Use of Proportion Reduction in Variance Statistics**

We have detailed throughout this chapter how the introduction of predictors in a hierarchical linear model can explain variance at both levels 1 and 2. We have also illustrated how to compute this proportion reduction in variance relative to some base model. In general, the principles introduced in this chapter for two-level organizational effects models extend directly to two-level growth models (see Chapter 6) and three-level models (see Chapter 8) as well. The use of these techniques, however, can become confusing and estimation anomalies can arise as the models grow more complex, especially at level 2 (or levels 2 and 3 in a three-level application.) For this reason,
we summarize below some key data analysis advice that has worked well on most problems that we have encountered.

**Random-Intercept-Only Models.** In these applications, we have two statistics to track, the proportion reduction in variance at level 1 and the proportion reduction in variance in the level-2 intercepts. Typically, the introduction of level-1 predictors will reduce the level-1 residual variance and may also change the level-2 variance, \( \tau_{00} \), as well. This occurs, in part, because as each new level-1 predictor is entered into the model, the meaning of the intercept, \( \beta_{0j} \), may change. (Recall that \( \beta_{0j} \) is the predicted outcome for an individual in unit \( j \) with the mean of 0 on all of the level-1 predictors.) As a result, \( \tau_{00} \) represents the variability for a different parameter. The only exception is when all level-1 predictors are group mean centered, in which case \( \beta_{0j} \) remains constant as the group mean. Formally, the introduction of predictors at level 1 need not reduce the level-2 variance. Indeed, the residual variance at level 2 can be smaller or larger than the unconditional variance.

This observation gives rise to a key principle in using the proportion reduction in variance statistics at level-2 in hierarchical models. Technically speaking, the variance explained in a level-2 parameter, such as \( \beta_{0j} \), is conditional on a fixed level-1 specification. As a result, proportion reduction in variance statistics at level 2 are interpretable only for the same level-1 model. Consequently, we recommend that researchers develop their level-1 model first, and then proceed to enter level-2 predictors into the analysis. Assuming the level-1 model remains fixed, no anomalies should arise in the computation of proportion reduction in variance as new level-2 predictors are entered into the equation for \( \beta_{0j} \). The proportion reduction in variance, or the “variance explained,” will increase as significant predictors enter the model. The introduction of nonsignificant predictors should have little or no impact on these “\( R^2 \)” statistics. We note that it is mathematically possible under maximum likelihood estimation for the residual variance to increase slightly if a truly nonsignificant predictor is entered into the equation. Such cases result in the computation of slightly negative-variance-explained statistics for the variable just entered. The negative differences here, however, will typically be quite small.

Snijders and Bosker (1999) offer an alternative approach to monitoring the proportion of variation explained. This approach, which applies to random intercept models, considers the “proportion reduction in prediction error” at each level; that is, the proportion reduction in error of prediction of \( Y_{ij} \) at level 1, and the proportion reduction in error of prediction of \( \bar{Y}_{.j} \) at level 2. This approach ensures that the proportion of variance explained is positive. Note that the variation explained is the variation in the observed outcome (which typically includes measurement error).

**Random-Slopes and -Intercepts Models.** Variance-explained statistics at level 2 can become more complicated when there are multiple random effects at level 2 as in the intercepts- and slopes-as-outcomes model for the social distribution of achievement illustrated above. These complications arise because the level-2 random effects may be correlated with one another. As a result, predictors entered into one level-2 equation can affect variance estimates in another equation.

The social distribution of achievement example was a relatively straightforward case because all of the level-1 predictors were group-mean centered and all of the coefficients were treated as random. (See the discussion below for further complexities that can arise when some level-1 predictors are grand-mean centered and/or their coefficients are fixed.) In cases like this, the only “anomaly” that may arise is that the introduction of a level-2 predictor in one equation can seemingly explain variance in another level-2 equation. This will occur when the errors in the random effects from the two equations are correlated and the predictor in question really belongs in both equations but is absent from one. For example, the introduction of a predictor for say the SES differentiation effect might result in a reduced residual variance estimate for the intercept term (i.e., the variability in school mean achievement) even though it had not yet been entered into the intercept equation. Such a phenomenon is evidence of a model misspecification, which can be corrected by entering the predictor in both equations. (A procedure to test for such model misspecification is detailed in Chapter 9.) We note that in extreme cases a predictor may appear to have a significant effect in one level-2 equation that totally disappears when the model misspecification is addressed in the other equation.

This gives rise to another model building principle for data analysis. Although it is possible to enter a different set of level-2 predictors for each outcome in the level-2 model, this flexibility should be used judiciously and with caution. The safest way to proceed is to introduce a common set of level-2 predictors in all of the level-2 equations. Unfortunately, following this advice is not always feasible as it can result in the estimation of an excessive number of fixed-effects parameters. In addition, in many applications a different variable set may be hypothesized to predict mean differences as compared to modeling differentiating effects. In these circumstances, we generally recommend specification of the intercept model, \( \beta_{0j} \), first. If additional predictors are considered for one or more of the slope equations, they should also be entered into the intercept model. Only if they are truly nonsignificant in the intercept model should they be deleted. Similarly, when two or more random slopes are moderately to highly correlated, any level-2 predictor entered into one of the equations should also be entered into the
other. Only if the effect is truly insignificant in one of the equations should it be dropped.

Complex Models with Fixed and Random Level-1 Coefficients and Mixed Forms of Centering. Anomalies in variance-explained statistics are most likely in these complex hierarchical models, and the reasons for this are not fully understood in all cases. In addition to following the advice already detailed above, we have generally found the following to work well. For each level-1 predictor, whether group- or grand-mean centered (or no centering), the aggregate of that level-1 predictor, \( X_{1..} \), should also be entered in the intercept model. This allows representation for each level-1 predictor of the two separate relationships, \( \beta_w \) and \( \beta_b \), that might exist. If \( \beta_w \) and \( \beta_b \) are different, as is often the case in organizational applications, the failure to represent both relations introduces a model misspecification. As noted above, the presence of model misspecifications can result in anomalous "explained-variance statistics" because of the correlated error terms at level 2. The safest procedure is to assume the coefficients are different and only restrict the model when evidence to the contrary has been assembled.

Estimating the Effects of Individual Organizations

Conceptualization of Organization Specific Effects

An important practical use of multilevel data is to monitor the performance of individual organizations—firms, schools, or classrooms, for example. One might use such data to hold organizational leaders accountable, to rank units for purposes of evaluation, or to identify for further study organizations that are unusually effective or ineffective.

We discuss below the uses of hierarchical models in this context. Our discussion is built around the problem of developing good individual school accountability indicators. The issues raised here, however, apply more generally.

Commonly Used Estimates of School Performance

The most common estimator of school performance uses the predicted mean outcome for each school based on the background and prior ability of students in that school. Schools that score higher than predicted are viewed as effective. Typically, the effectiveness score or performance indicator is just the actual mean achievement minus the predicted mean score for each school. Specifically, in the case of one background variable, the performance indicator or school effect is

\[
Y_{i..} - \bar{Y}_{..} - \beta(X_{i..} - \bar{X}_{..}),
\]

where

- \( Y_{i..} \) is the mean outcome for school \( i \);
- \( \bar{Y}_{..} \) is the grand-mean outcome across all schools;
- \( \bar{X}_{i..} \) is the school mean on the background variable \( X_{ij} \);
- \( \bar{X}_{..} \) is the grand mean of \( X \); and
- \( \beta \) is a coefficient of adjustment.

A focus of the methodological controversy has been the choice of method for estimating \( \beta \). Aitkin and Longford (1986) summarize common alternatives. One of their key conclusions is that estimates of \( \beta \) based on OLS regression (ignoring students' membership in schools) can be quite misleading. An alternative method is to use an ANCOVA model in which school membership is included as a series of dummy variables. However, this approach quickly becomes impractical as the number of schools increases. Both OLS regression (ignoring group membership) and the ANCOVA can produce unstable estimates of school effects when sample sizes per school are small. Raudenbush and Willms (1995) discuss threats to valid inference in using these indicators.

Use of Empirical Bayes Estimators

The empirical Bayes residuals estimated under a hierarchical linear model provide a stable indicator for judging individual school performance. These empirical Bayes estimates have distinct advantages over previous methods. They (a) take into account group membership even when the number of groups is large and (b) produce relatively stable estimates even when sample sizes per school are modest.

The random-intercepts model discussed earlier is a common choice for those applications. At the student level,

\[
Y_{ij} = \beta_{0j} + \sum q \beta_{qj} (X_{qij} - \bar{X}_{q..}) + r_{ij},
\]

At the school level,

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

and

\[
\beta_{qj} = \gamma_{q0} \quad \text{for } q = 1, \ldots, Q.
\]
where the $\gamma_{00}$ is the overall intercept. Notice that all level-1 regression coefficients except the intercept are constrained to be constant across schools, and the unique individual-school effect is just $u_{0j}$.

In the combined form, the model is

$$Y_{ij} = \gamma_0 + \sum_q \gamma_q (X_{qij} - \bar{X}_{q..}) + u_{0j} + r_{ij}. \quad [5.55]$$

This model hypothesizes that all students within school $j$ have an effect, $u_{0j}$, added to their expected score as a result of attending that school. Formally, this is an ANCOVA model where the $X$s are covariates and the set of $J$ schools constitute independent groups in a one-way random-effects ANCOVA. A key difference from a traditional random-effects ANCOVA, however, is that our goal here is literally to estimate the effects for each level-2 unit and not just their variance.

_Estimation._ An OLS estimator of the effect for each school $J$ is

$$\hat{u}_{0j} = \bar{y}_{.,j} - \hat{\gamma}_{00} - \sum_q \hat{\gamma}_q (\bar{X}_{q.,j} - \bar{X}_{q..}). \quad [5.56]$$

Note that $\hat{u}_{0j}$ in Equation 5.56 is just the school mean residual for the ANCOVA. Schools with the small samples, $n_j$, will tend to yield unstable estimates for $\hat{u}_{0j}$. As a result, these schools are more likely to appear extreme purely as a result of chance. Thus, to select out the largest or smallest values of $\hat{u}_{0j}$ as indicators of the most or least effective organizations is likely to capitalize on extreme chance occurrences.

As noted in Chapter 4, the estimator for $u_{0j}$ in the hierarchical analysis, $u^*_{0j}$, is an empirical Bayes or shrinkage estimator. The $u^*_{0j}$ shrinks the OLS school-effect estimate proportional to the unreliability of $\hat{u}_{0j}$. The more reliable the OLS estimate from a school is, the less shrinkage occurs. Hence, values that are extreme because they are unstable will be pulled toward zero.

Specifically, the estimated school effect under a hierarchical analysis is

$$u^*_{0j} = \lambda_j \hat{u}_{0j}, \quad [5.57]$$

where

$$\lambda_j = \tau_{00}/(\tau_{00} + \sigma^2/n_j)$$

is a measure of the reliability of $\hat{u}_{0j}$ as an estimate of $u_{0j}$.

**Threats to Valid Inference Regarding Performance Indicators**

Despite the technical advantages of empirical Bayes estimators, major validity issues remain in using such statistics as performance indicators. We consider some of these below. See Randenbush and Wills (1995) for a more detailed discussion.

Figure 5.5. A Hypothetical Example of the Relationship between Achievement and Student Social Class in Two Groups of Schools: Low-Social-Class Schools (Group 1) and High-Social-Class Schools (Group 2)

_Bias._ Studies of performance indicators may profitably be viewed as quasi-experiments in which each unit is a treatment group. The problem of valid causal inference in such settings has been extensively studied (Cook & Campbell, 1979). When random assignment of subjects to treatments is impossible, an attempt must be made to identify and control for individual background differences that are related to group membership and also to the outcome. This poses two problems: First, one can never be confident that all of the relevant background variables have been identified and controlled. Second, reasonable people can disagree about proper models for computing adjustment coefficients, and this choice of adjustments can have a substantial impact on inferences about the individual school effects. One general principle does emerge, however, in considering adjustments: The more dramatically different the groups are on background characteristics, the more sensitive inferences are likely to be to different methods of adjustment and the less credible the resulting inferences.

This principle can be illustrated by a deliberately exaggerated and hypothetical example. Consider Figure 5.5. Student achievement (vertical axis) is plotted against student SES (horizontal axis). Associated with each school is a regression line describing the relationship between social class and achievement. The length of each line represents the range of social class within each
school. Notice that there are two types of schools: The students in Group 1 schools are of low SES, and in Group 2 schools, SES is considerably higher. Within Groups 1 and 2, schools vary only slightly in their effectiveness, as indicated by the small distances between the parallel regression lines. Notice also that there is a compositional effect, D. That is, Group 1 and Group 2 schools differ by more than one would predict given the regression of student achievement on individual SES, and this appears related to the average SES of the schools.

Now we shall consider two alternative models for estimating the individual school effects. The first is a fixed-effects ANCOVA controlling for student social class but ignoring the compositional effect. Formally, using the model implied by Equation 5.51,

$$Y_{ij} = \beta_{0j} + \beta_1(X_{ij} - \bar{X}.\) + r_{ij},$$  

[5.58]

$\beta_1$ is the adjustment coefficient associated with student SES and $\beta_{0j}$ represents the fixed school effect in this case. Because the achievement-SES relationship is homogeneous across the six schools, the difference in effects between any two schools is the distance between their regression lines. For example, the difference between School 2 and School 5 is very large. Notice that under this model every Group 2 school will appear more effective than any Group 1 school.

The second model explicitly controls for school SES when estimating the school effects. This involves extending Equation 5.58 to include mean school SES, $\bar{X}_{.,j}$, as a second covariate. For example,

$$Y_{ij} = \beta_{0j} + \beta_1(X_{ij} - \bar{X}_{.,j}) + \beta_2(\bar{X}_{.,j} - \bar{X}.\) + r_{ij},$$  

[5.59]

In practice, Equation 5.59 implies comparing "like with like." Now Group 1 schools are compared to each other, and Group 2 schools are compared to each other. From Figure 5.5, we see that Schools 2 and 5 are now viewed as equally effective: Each is about average compared with schools in its group (schools having similar social composition).

This strategy may seem fairer than the first, but is it? Suppose that the Group 2 schools have more effective staff and that staff quality, not student composition, causes the elevated test scores. The results in Figure 5.5 could occur, for example, if the school district assigned its best principals and teachers to the more affluent schools. If so, the second strategy would give no credit to these leaders for their effective practices.

The key concern is that without having formulated an explicit model of school quality, we can never be sure that we have disentangled the effect of school composition from other school factors with which composition is often correlated.

![Figure 5.6. Identifying School Effects when the Relationship between Achievement and Student Background is Heterogenous](image)

**Heterogeneity of Regression.** Problematic as this may seem, the example above is far simpler than reality is likely to be. Our example assumes that the SES-achievement regression lines are identical in all schools. In many cases, the regressions may be heterogeneous, as indicated in Figure 5.6. Now, regardless of the method of adjustment, the estimate of a school's effectiveness will depend on the social class of the child in question. For example, School 1 appears very effective relative to School 2 for low-SES students (i.e., effect 1 in Figure 5.6). However, for high-SES students, the differences between two schools are negligible (i.e., effect 2).

**Shrinkage as a Self-Fulfilling Prophecy.** As discussed in Chapter 3, the shrinkage estimator, $\hat{u}_{0j}$, has a smaller expected mean-squared error than does the least squares estimator, $\hat{u}_{0j}$, and protects us against capitalizing on chance. However, shrinkage estimators are conditionally biased. From Equation 5.56, we can see that the expected value of $\hat{u}_{0j}$ given the true value of $u_{0j}$ is

$$E(\hat{u}_{0j} | u_{0j}) = u_{0j} = \mu_{rj} - \sum_q \gamma_q (\bar{X}_{q..} - \bar{X}_{..}),$$  

[5.60]
where $\mu_{y_j}$ is the unadjusted mean outcome in group $j$. Notice that $u_{0j}$ is the deviation of group $j$'s unadjusted mean, $\mu_{y_j}$, from a value predicted on the basis of student-background variables. The conditional expectation of the empirical Bayes estimator, $u_{0j}^{*}$, is

$$E(u_{0j}^{*} | u_{0j}) = \lambda_j \mu_{0j},$$

so the bias is

$$\text{bias}(u_{0j}^{*} | u_{0j}) = -(1 - \lambda_j) \left[ \mu_{y_j} - \sum_q \gamma_{qj} (\bar{X}_{q-j} - \bar{X}_{q}) \right].$$

This formula indicates that to the extent $\hat{u}_{0j}$ is unreliable, the estimate $u_{0j}^{*}$ will be biased toward the value predicted on the basis of student background. For example, unusually effective schools that have children of disadvantaged backgrounds will have their high mean effect estimates biased downward toward the value typically displayed by other, similarly disadvantaged schools. This procedure then operates as a kind of statistical self-fulfilling prophecy in which, to the extent the data are unreliable, schools effects are made to conform more to expectations than they do in actuality.

Power Considerations in Designing Two-Level Organization Effects Studies

A key consideration in designing two-level studies of organization effects is choosing the sample size, $n$, per organization as well as the number of organizations, $J$. The best choice of $n$ and $J$ depends strongly on the aims of the study. We briefly consider three cases: where the key explanatory variable of interest is at level 2; where the key explanatory variable is at level 1; and where primary interest focuses on a cross-level interaction (i.e., the effect of a level-2 predictor on a level-1 slope coefficient).

Key Explanatory Variables at Level 2. An optimal $n$ per organization depends on the cost of sampling units at each level and the variability at each level. A large $n$ is helpful when (a) it is comparatively expensive to sample organizations (as compared to sampling organization members once an organization has been selected) and (b) variation within organizations is comparatively large (Raudenbush, 1997; Snijders & Bosker, 1999). In the simple case of a balanced design with one level-1 predictor and a single level-2 predictor, a well-known result from sampling research (cf. Cochran, 1975) is that the optimal $n$ per cluster is

$$n_{\text{opt}} = \left( \frac{C_2}{C_1} \frac{1}{\rho} \right)^{1/2}, \quad \rho > 0,$$

where $C_2/C_1$ is the relative cost of sampling at level 2 versus level 1 and $\rho = \tau_{00}/(\tau_{00} + \sigma^2)$ is the intra-organization correlation. Raudenbush (1997) discusses how adding covariates at level 1 can increase power by reducing $\tau_{00}$ and $\sigma^2$.

Key Explanatory Variable at Level 1. When the key explanatory variable is at level 1, interest may focus on the average effect of the level-1 variable, that is, $\gamma_{10}$. Suppose, for example, the level-1 variable is an indicator for a treatment ($X_{ij} = 1$ if experimental, 0 if control) and there are no other predictors at any level. Then the optimal $n$ per organization

$$n_{\text{opt}} = \sqrt{\frac{C_2}{C_1} \frac{\tau_{11}}{\tau_{11}}}, \quad \tau_{11} > 0,$$

where $\tau_{11}$ is the variance of the level-1 slope and $\sigma^2$ is the level-1 or “within-cluster” variance. We note that if the primary aim is to estimate $\tau_{11}$ with precision the required $n$ will typically be larger than $n_{\text{opt}}$ (Raudenbush & Liu, 2000).

Cross-level Interactions. Suppose the aim is to make inferences about a cross-level interaction; for example, a treatment indicator at level 2, $W$, affects a within-unit slope associated with some level-1 predictor, $X_j$. Assuming $\tau_{11} > 0$ (where $\tau_{11}$ is the residual variance in the level-1 slope after controlling for $W_j$), Equation 5.64 will apply. However, if $\tau_{11} = 0$, precision will depend on the total sample size $nJ$.

A Windows-based computer program to calculate optimal sample designs and power for a variety of organizational effect and growth study designs has been developed by Raudenbush and Liu and is available at www.sscicentral.com.