Chapter 1 Introduction to Multilevel Models

1.1 Nested Data Structures

Objectives

- Understand two basic types of nested data structures, hierarchical and longitudinal.
- Discern sources of nesting among observations, intended or unintended.

3

Nested Data

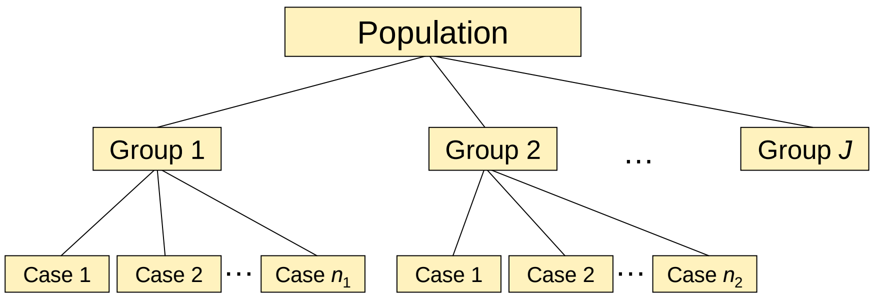
- Multilevel models are employed when we have nested data.
- Nested data typically comes in one of two forms, hierarchical or longitudinal.

4

A term you will hear often when working with multilevel models is *nested data* or, more simply, *nesting*. Two common sources of nested data are hierarchical and longitudinal data structures.

Hierarchical Data Structures

- Hierarchical data structures are those in which multiple micro-level units are sampled for each macro-level unit.
- A common hierarchical data structure is when individuals (micro-units) are sampled from naturally occurring groups (macro-units).



5

In hierarchical data structures, there are (at least) two levels of sampling. Macro-units (for example, groups) are sampled and then multiple micro-units (for example, individuals) are sampled within each macro-unit.

We will typically refer to these two levels of sampling as *Level 1* and *Level 2*, respectively.

A common example of hierarchically structured data comes from the education field: students are nested within classrooms, which in turn might be nested within schools. Another example from medical research is that multiple patients may be seen by the same physician.

Other examples of hierarchically structured data include these:

- siblings nested within families
- families nested within neighborhoods
- employees nested within managers
- managers nested within sales districts
- tool operators nested within machines

Unintentional Sources of Nesting

- Nesting might also occur even when not an explicit part of the study design, and thus be unintentional.
- Examples include
  - respondents nested within interviewer
  - homeless adolescents nested within social service sectors
  - multiple specimens nested within laboratory

6

Nesting might also occur even when not an explicit part of the study design. This is sometimes called *unintentional nesting*. Here are some examples:

- A random sample of 500 individuals might be drawn from a population and assumed to be independent. However, each of these individuals is interviewed by one of 10 interviewers. Many of the questions address sensitive topics, and some interviewers are better skilled at building rapport with the respondents than are others. It is possible that the individuals who share the same interviewer are more similar to one another than to subjects assessed by another interviewer.
- A random sample of homeless adolescents are recruited from public gathering places in a large metropolitan city. The study is designed so that an independent sample is obtained. However, adolescents congregate to different parts of the city as a function of what type of social services they value (quality of food versus sleeping facilities, etc.). It is possible that the individuals who self-select to parts of the city based on social services are more similar to one another than to individuals in other parts of the city.
- It is not uncommon to pool data collected across multiple laboratories using the same experimental procedures. These studies are designed so that all lab specimens are independent given the shared experimental methodologies. However, small differences in procedures unique to each laboratory may induce non-independence given that specimens from a given lab are more similar to one another than are specimens from different labs.

Dependence in Hierarchical Data

- Because many micro-level observations come from the same macro-level unit, this produces dependence in the data.
  - Students attending the same school may have more similar academic outcomes than students attending different schools.
  - Employees working under the same manager may have more similar problem-solving strategies than employees working under different managers.
- Multilevel models provide a way to model this dependence, whereas more traditional models do not.

7

Hierarchical data structures pose challenges to traditional methods of data analysis because they produce *dependence*. That is, observations from micro-units that share the same macro-unit are expected to be correlated. For instance, students attending the same school may have more similar academic outcomes than students attending different schools – the achievement scores of students within a school would then be correlated.

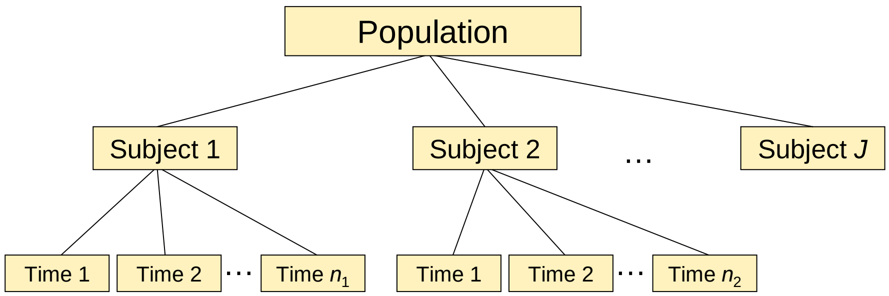
Most traditional data analysis models (for example, ANOVA, regression) instead assume independence. That is, they require that knowing one student's achievement score tells you nothing about any other student's achievement score. If students are nested within schools, however, this is unlikely to be the case. High-performing students may be clustered together at the best schools.

As we will see, multilevel models were explicitly developed to take account of dependence due to hierarchical data structures and provide a number of advantages over more traditional data analysis methods. For instance, in the example above, multilevel models would allow you to examine the effects of student-level variables on achievement (for example, SES), school-level variables (for example, private or public sector schools), and interactions between these two levels. Often, such person × context interactions are of key interest. As such, we will begin our foray into multilevel modeling by focusing on this data type.

There are a variety of terms used to describe these statistical models. These are known as variance components models, multilevel models, random coefficients models, random-effects models, mixed models, general mixed models, and hierarchical linear models. We use the term multilevel models throughout this course.

Longitudinal Data Structures

- Longitudinal data structures arise when the same units are sampled repeatedly over time.
- Longitudinal data is useful for tracking change over time in an outcome (for example, response to a drug).



8

In longitudinal data structures, repeated measurements are made on the same individual over time with the goal of tracking the progress or change on outcomes of interest. We can also think of longitudinal data structures as having two levels of sampling: Participants are sampled and then time-sampling of the repeated measures is performed for each person. Often, such assessments are performed on a specified fixed schedule, such as annually.

Like the hierarchical data structure, we will again typically refer to these as *Level 1* and *Level 2*, respectively.

In much longitudinal data, the behavior or health of people is tracked over time, but longitudinal data can also be used to track performance trends for companies or changes in life expectancies for countries, among other possibilities.

Dependence in Longitudinal Data

- Because repeated measures are collected on the same unit, this produces dependence in the data.
- Example: Employee job performance is tracked over a period of four years.
  - Some employees perform at consistently higher levels compared to other employees.
  - Some employees increase in performance at a steeper rate over time compared to other employees.
- Again, multilevel models provide a way to model this dependence, whereas more traditional models do not

9

Longitudinal data structures also produce *dependence*. In this case, we expect that individuals differ systematically from one another, producing within-person correlations on the repeated measures.

For instance, if the job performance of employees is tracked annually, it can be expected that some employees will perform consistently above average, while others will perform consistently below average. Similarly, if you were to study changes in weight due to the implementation of a weight-loss program, it could be expected that some people would be generally heavier than others at all time points due to their height and build.

The differences between people need not only be differences in level. There could also be differences in patterns or rates of change across individuals. For instance, changes in marital satisfaction following the birth of a child may differ across couples in predictable ways.

This dependence once again poses a problem to traditional data analysis models. Models that assume independence require that knowing an individual's value on the outcome at one point in time tells you nothing about their value at the next point in time, a highly unlikely situation.

Fortunately, multilevel models provide a way to deal with dependence in longitudinal data structures. Moreover, these models allow us to examine across-person heterogeneity in change over time, and predictors of change over time. We will thus turn to multilevel models for longitudinal data as the second primary topic of this course.

First, however, we examine the potential consequences of ignoring the nesting in the data by using traditional regression methods to analyze a set of data. This will also permit us to review the standard regression model, the understanding of which will be critical for extending to the multilevel model.

1.2 Ignoring Dependence

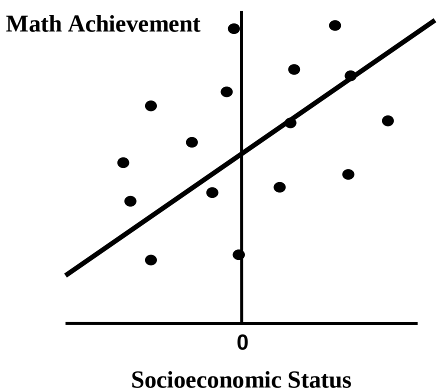
Objectives

- Review the standard single-level regression model.
- Introduce the HSB data set.
- Clarify why standard statistical techniques are inappropriate for nested data structures.

11

A Standard Analysis

- Suppose that we are interested in evaluating the effect of socioeconomic status on math achievement in high school students.
- We might begin by fitting a simple linear model:



$$y_i = \beta_0 + \beta_1 x_i + r_i$$
$$\text{Math}_i = \beta_0 + \beta_1 \text{SES}_i + r_i$$

12

To better understand the consequences of dependence in the data, we will begin with an analysis that ignores it. For our example, we are concerned with the relationship between socioeconomic status and math achievement.

Here we have displayed the scatter plot and simple linear fixed effects regression of math achievement on socioeconomic status. Socioeconomic status has been centered so that zero is average SES, negative values indicate below average SES, and positive values indicate above average SES. The equation for a generic simple regression model is

$$y_i = \beta_0 + \beta_1 x_i + r_i$$

y_i The observed value of the dependent variable for individual *i*

x_i The observed value of the independent variable/predictor/covariate for individual *i*

β_0 The *intercept* of the regression line

β_1 The *slope* of the regression line, indicating the expected change in *y* per one-unit change in *x*

r_i The residual of the observed value y_i from the regression line $\beta_0 + \beta_1 x_i$ for individual *i*

In the present example, math achievement scores are the dependent variable and SES is the predictor. Our primary interest is in the slope estimate, indicating the magnitude of the effect of SES on math achievement.

The regression line itself indicates the expected value of *y* given *x*, or

$$E(y \mid x) = \beta_0 + \beta_1 x$$

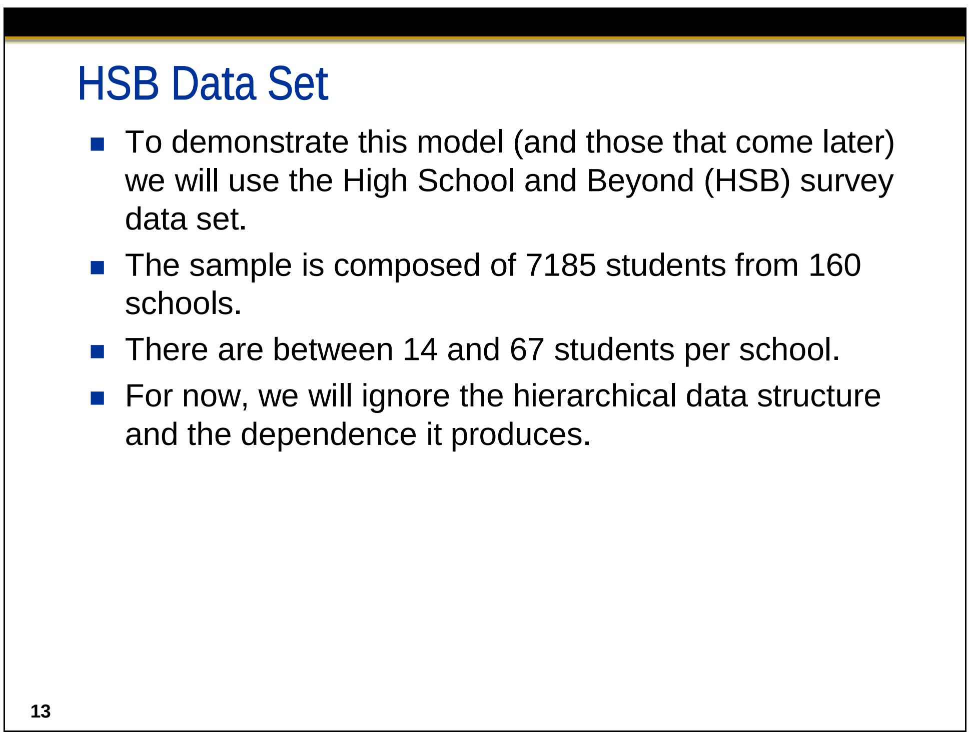
$E(\bullet)$ The expected value operator, returning the long-run (conditional) average of a random variable, or population (conditional) mean

Clearly, for a student with a value of zero on *x*, the expected value for *y* would simply be $E(y \mid x = 0) = \beta_0$. This, then, is our interpretation of the intercept. Of course, the intercept is only useful to interpret if zero is within the observed range of *x*. In our case, by centering SES so that its mean is zero, β_0 is interpretable as the predicted math achievement of a high school student of average SES.

The interpretation of the slope, β_1, is that it is the expected change in *y* per unit change in *x*. We can see this by comparing the expected value, where $x = 0$ to the expected value where $x = 1$.

$$E(y \mid x = 0) = \beta_0$$
$$E(y \mid x = 1) = \beta_0 + \beta_1$$

Thus β_1 is the expected change in *y* as *x* increases 1 unit. More simply, it is the *rise over run*. This is the parameter of primary interest in our example: we want to know how much math achievement improves as we move up the socioeconomic ladder.



To review the standard regression model, we will fit it to a real data set, the High School and Beyond public use data set. As noted above, this data set is actually hierarchically structured, but for the time being we will ignore that feature of the data.

HSB Simple Regression

c1_hsb_simplereg.sas

For this demonstration and many of those that follow we will use the HSB data set that was originally compiled by Stephen Raudenbush and Anthony Bryk for their book *Hierarchical Linear Models*. The data set **`mixed.hsb`** contains data on 7185 students from 160 schools on the following variables:

`student_id`	numeric code that uniquely identifies each student
`student_mathach`	continuous variable that indicates mathematical ability
`student_ses`	grand mean centered continuous variable that indicates the student's socioeconomic status
`student_min`	binary variable that indicates whether a student is a minority
`student_female`	binary variable that indicates whether the student is female
`school_id`	numeric code that identifies the school the student attends
`school_size`	number of students attending that school
`school_sector`	binary variable that indicates if the school is public (coded 0) or private (coded 1)
`school_disclim`	continuous variable that measures the disciplinary climate of the school

Note that there are two ID variables, one that uniquely identifies each student, **`student_id`**, and one that uniquely identifies each school, **`school_id`**. Further, some variables vary across students (even within a school), whereas others vary only across schools. To differentiate these two types of variables, we adopt the naming convention **`student_x`** for any student-level variable and **`school_w`** for any school-level variable. Making these distinctions will become critically important when we move to the multilevel model.

For now, however, we will concern ourselves only with the prediction of math achievement by the socioeconomic status of the student, ignoring the fact that there is nesting of students within schools.

We next examine the partial contents of the data set using PROC PRINT.

```
proc print data=mixed.hsb;
   var school_id student_id student_ses school_disclim
        student_mathach;
   where student_id le 75;
run;
```

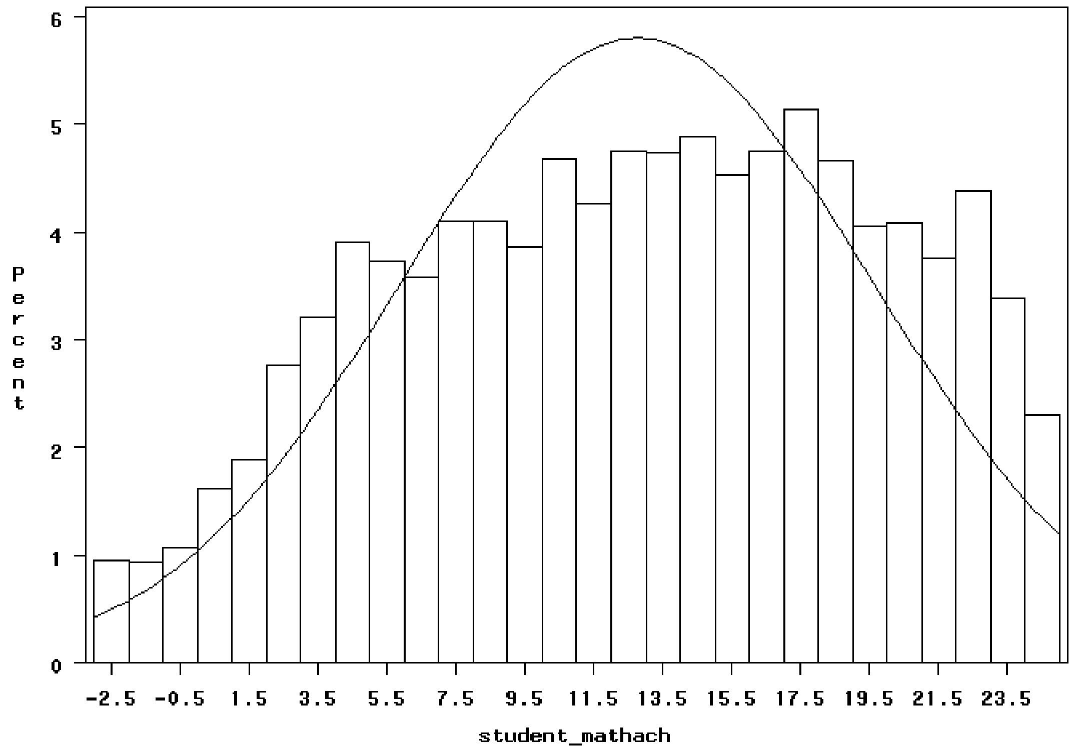
Obs	school_ ID	student_ ID	student_ ses	school_ disclim	student_ mathach
1	1224	1	-1.528	1.597	5.876
2	1224	2	-0.588	1.597	19.708
3	1224	3	-0.528	1.597	20.349
4	1224	4	-0.668	1.597	8.781
5	1224	5	-0.158	1.597	17.898
<snip>					
45	1224	45	0.752	1.597	23.584
46	1224	46	0.012	1.597	14.053
47	1224	47	-0.418	1.597	2.183
48	1288	48	-0.788	0.174	7.857
49	1288	49	-0.328	0.174	10.171
<snip>					
71	1288	71	0.322	0.174	23.578
72	1288	72	0.592	0.174	12.810
73	1296	73	-0.148	-0.137	12.668
74	1296	74	-0.858	-0.137	0.781
75	1296	75	-1.568	-0.137	4.968

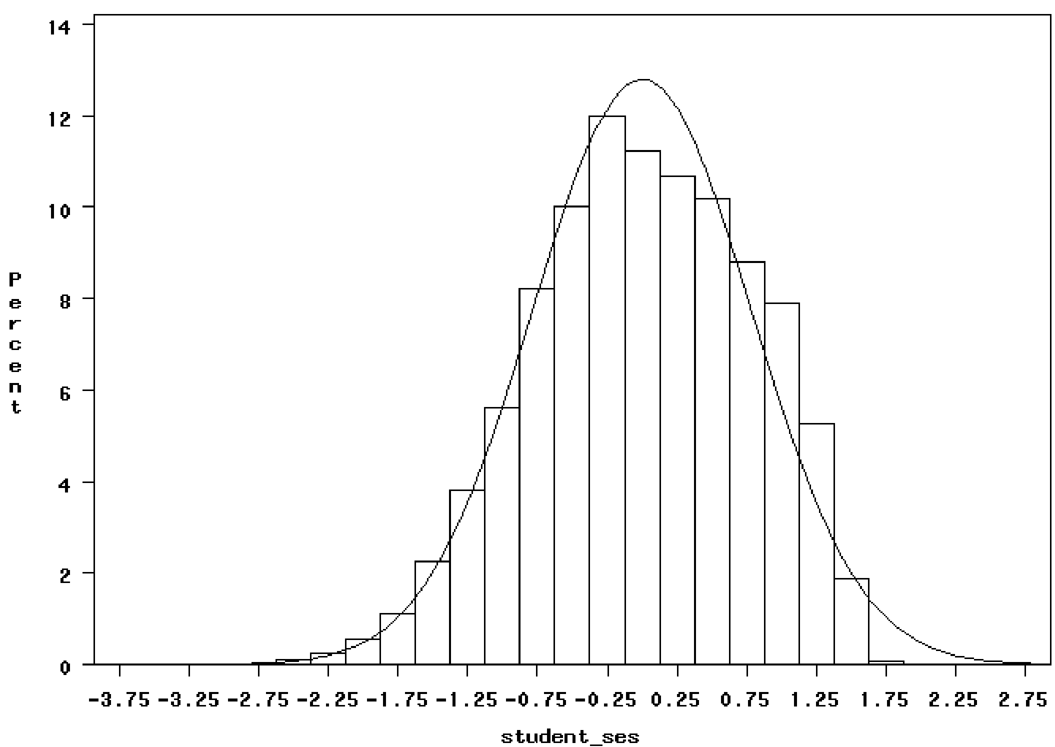
These results show the inherent nesting of the data in that there are multiple students nested within schools; further, the student-level variables vary across both students and schools, but the school-level variables are constant across student within school, but vary across schools.

Examining Univariate and Bivariate Distributions

We can next examine histograms for the predictor **`student_ses`** and the outcome **`student_mathach`**.

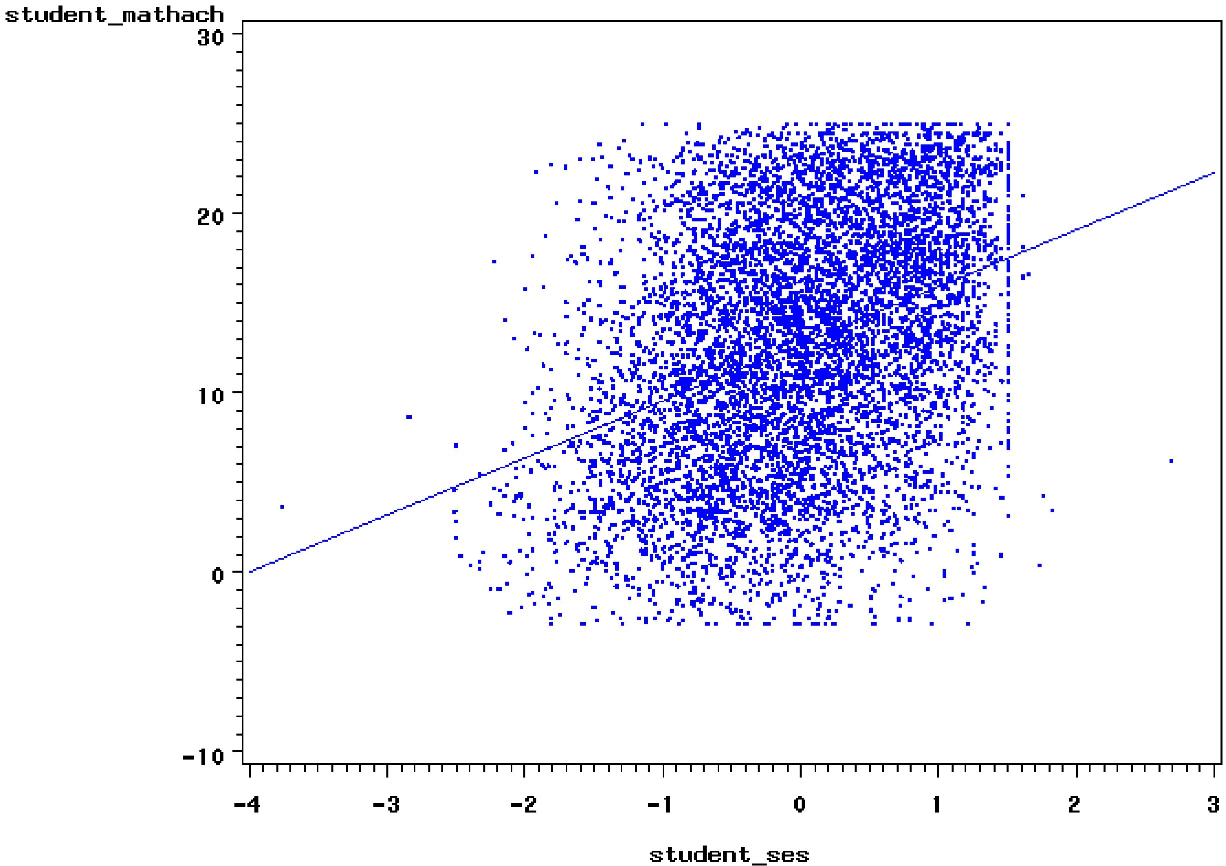
```
goptions reset=all hsize=7 vsize=5;
proc univariate data=mixed.hsb;
   var student_mathach student_ses;
   histogram/normal(color=black);
run;
```


Percent
0
1
2
3
4
5
6
-2.5 -0.5 1.5 3.5 5.5 7.5 9.5 11.5 13.5 15.5 17.5 19.5 21.5 23.5
student_mathach


Percent
0
2
4
6
8
10
12
14
-3.75 -3.25 -2.75 -2.25 -1.75 -1.25 -0.75 -0.25 0.25 0.75 1.25 1.75 2.25 2.75
student_ses

In addition to the univariate histograms, we can examine the bivariate relationship between student SES and student math achievement.

```
goptions reset=all hsize=7 vsize=5;
proc gplot data=mixed.hsb;
   plot student_mathach*student_ses;
   symbol v=dot w=.1 h=.1 i=rl l=1 c=blue;
run; quit;
```



The plot shows a modest, positive effect of SES on math achievement. This also reveals several interesting characteristics about the data.

First, there appears to be some limits on the measurement of both scales (as is reflected in the sharp lines defining three of the four sides of the scatter plot).

Second, there are a small number of observations that stray beyond the apparent upper limit of student SES; it might be of use to explore these cases further in later analyses.

Univariate Regression Model

```
proc reg data=mixed.hsb;
   model student_mathach = student_ses / clb stb;
run;quit;
```

Selected REG procedure statements

MODEL names the dependent variable (criterion) to the left of the equal sign and independent variables (predictors) to the right of the equal sign. In this MODEL statement, **`student_mathach`** is to be predicted by **`student_ses`**.

The specified MODEL statement options are as follows:

CLB requests 1-α confidence intervals for the estimates (default α = .05).

STB requests the standardized regression coefficient estimates (unstandardized coefficients are provided by default).

PROC REG Output:

```
                              The REG Procedure
                                Model: MODEL1
                    Dependent Variable: student_mathach

                  Number of Observations Read        7185
                  Number of Observations Used        7185

                           Analysis of Variance

                                   Sum of          Mean
 Source                  DF       Squares        Square   F Value   Pr > F

 Model                    1         44233         44233   1074.69   <.0001
 Error                 7183        295644      41.15882
 Corrected Total       7184        339877

           Root MSE              6.41551    R-Square     0.1301
           Dependent Mean       12.74785    Adj R-Sq     0.1300
           Coeff Var            50.32623

                           Parameter Estimates

                           Parameter       Standard
Variable           DF       Estimate          Error    t Value    Pr > |t|

Intercept           1       12.74740        0.07569     168.42      <.0001
student_ses         1        3.18387        0.09712      32.78      <.0001

                            Parameter Estimates

                             Standardized
    Variable           DF        Estimate       95% Confidence Limits

    Intercept           1               0       12.59903       12.89576
    student_ses         1         0.36076        2.99348        3.37426
```

- First, we see that our model is explaining very little. The R^2 for the model is only .13, indicating that socioeconomic status explains only 13% of the variance in math achievement scores.
- The intercept estimate of 12.7 is the expected math achievement score of a student of average SES. This is only of modest interest.
- More importantly, the slope estimate of 3.2 indicates that, on average, math achievement scores increase 3.2 units per one unit increase in SES. The test of this estimate is statistically significant.
- The 95% confidence interval for the SES effect ranges from 3.0 to 3.4, indicating that we are estimating this effect with a good deal of precision.
- The standardized estimate for student SES of .36 reflects that, on average, a one standard deviation increase in student SES is associated with a .36 standard deviation increase in student math achievement.

Of course this is an overly simplistic model in that we are only considering a single predictor of student math achievement. In most any applied setting, we would expand this model to include multiple predictors.

Multiple Regression

- We might then want to add a second predictor, whether the student attends a school in the private or public sector.
- The new model equation would be

$$\text{Math}_i = \beta_0 + \beta_1\text{SES}_i + \beta_2\text{Sector}_i + r_i$$

- Note that β_1 and β_2 are now partial regression coefficients. These coefficients indicate the effect of each predictor holding constant (controlling for) the other.

15

Given that we have only explained 13% of the variance, we might want to add a second predictor to the model to better understand individual differences in math achievement. For instance, we might believe that students attending private high schools will out-perform those in public high schools.

Adding another predictor to our regression model moves us out of simple regression and into multiple regression. The interpretation of the model coefficients is correspondingly more complex.

A general multiple regression model can be written as

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + r_i$$

- We now have p + 1 regression parameters to estimate, the intercept plus one slope for each predictor.
- The expected value for y is now $E(y \mid x_1, x_2, \ldots, x_p)$, or the optimal linear combination of the set of predictors. We will use the notation $\hat{y}$ for a shorthand expression of this, so the model implies that

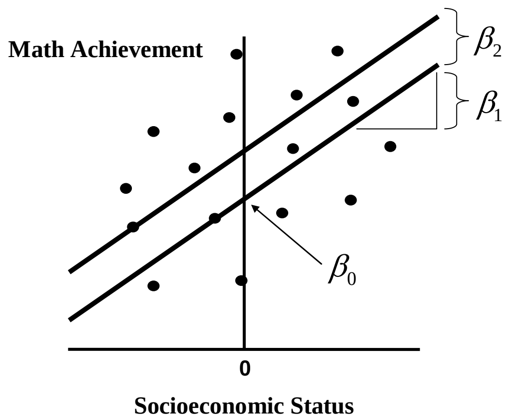
$$\hat{y}_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} .$$

- The R^2 for the model now represents the variability in the outcome variable explained by the set of all predictors, or the squared correlation between y_i and $\hat{y}_i$.
- The parameter estimates indicate the effect of each predictor holding constant (net of, controlling for, adjusting for) the influence of the other predictors. They are thus sometimes referred to as *partial regression coefficients*. A partial regression coefficient still represents the expected change in y given a one unit change in the predictor, but now we must add "holding all other predictors constant".
- For instance, in our example, β_2 represents the effect of school sector holding SES constant. What this means is that β_2 is the expected achievement difference between students attending private versus public schools who are of the same socioeconomic status.

Interpretations in Multiple Regression

- We might expect the resulting model to look something like this:

$$\text{Math}_i = \beta_0 + \beta_1\text{SES}_i + \beta_2\text{Sector}_i + r_i$$



16

This particular model is of the same form as an analysis of covariance (ANCOVA). Though students of public and private schools may differ in their average socioeconomic status, the estimated effect for sector adjusts for this difference.

This adjustment is only possible because the model assumes that the effect of sector on a student's math achievement is the same regardless of their level of SES. This is the ANCOVA assumption of parallel regression lines.

If the assumption of parallel regression lines is not true, then the results of the model can be misleading and a better model would include an interaction between SES and Sector. We will explore the estimation and interpretation of interactions in depth in Chapter 3.

HSB Multiple Regression

c1_hsb_multireg.sas

We will now augment our earlier regression model with a second predictor, **school_sector**.

```
proc reg data=mixed.hsb;
   model student_mathach = student_ses school_sector/ clb stb;
run;quit;
```

PROC REG Output:

```
                            The REG Procedure
                              Model: MODEL1
                   Dependent Variable: student_mathach

                  Number of Observations Read        7185
                  Number of Observations Used        7185

                          Analysis of Variance

                                  Sum of         Mean
 Source                  DF      Squares       Square   F Value   Pr > F

 Model                    2        50716        25358    629.83   <.0001
 Error                 7182       289161     40.26191
 Corrected Total       7184       339877

           Root MSE             6.34523    R-Square     0.1492
           Dependent Mean      12.74785    Adj R-Sq     0.1490
           Coeff Var           49.77487

                          Parameter Estimates

                          Parameter      Standard
Variable           DF      Estimate         Error    t Value    Pr > |t|

Intercept           1      11.79325       0.10610     111.15      <.0001
student_ses         1       2.94856       0.09783      30.14      <.0001
school_sector       1       1.93501       0.15249      12.69      <.0001

                           Parameter Estimates

                            Standardized
    Variable           DF        Estimate       95% Confidence Limits

    Intercept           1               0       11.58526       12.00125
    student_ses         1         0.33409        2.75678        3.14033
    school_sector       1         0.14066        1.63608        2.23394
```

Note that our model R^2 has increased slightly. Despite this, the effect of private versus public schooling is still statistically significant (owing in part to very high power). The effect of SES is still positive, but is somewhat diminished by controlling for sector. The expected achievement difference between students of the same SES in private versus public schools is 1.9, with a 95% confidence interval from 1.6 to 2.2.

Assumptions of the Regression Model

- The ordinary regression model makes a number of assumptions, including:
  - sum (and mean) of residuals is zero
  - predictors uncorrelated with residuals
  - predictor levels are fixed (not random or sampled)
  - residuals are normally distributed (for inference)
  - residuals are homoscedastic (have constant variance)
  - residuals are independent (uncorrelated)
    - These last three are reflected in the term *normal-iid* meaning that the residuals are normal, independent, and identically distributed.

18

The assumptions of the ordinary least-squares regression model are these:

1. The sum of the residuals is zero (implying the mean is also zero). What this means is that if the observed values for *y* are systematically higher or lower than they ought to be, our regression model will not be able to account for this. For instance, if *y* is weight and we use an improperly calibrated scale that measures everyone as 5 pounds too heavy, then this will bias the intercept.

2. The predictors are uncorrelated with the residuals. If the predictors are correlated with the residuals, then this will bias the estimation of the effect of the predictor. This can arise, for instance, if there are important predictors of *y* that have been omitted from the model that correlate with the predictors that are included in the model. The result is referred to as *omitted variable bias*.

3. The predictor levels are fixed, not random. This assumption implies that the researcher selected the levels of the *x* variables at which to observe *y*. Outside of experimental designs, this is rarely the case. Typically, we observe *x* simultaneously with *y* and both are random variables. Fortunately, treating *x* as fixed (even though it is random) only results in a modest loss of efficiency and power but leaves the regression coefficients unbiased.

4. Residuals are normally distributed. This assumption is required only for inference. If the residuals are normally distributed, then the *F*-statistic for the regression model and the *t*-tests of the regression coefficients are exact. If the residuals are not normally distributed, then this is no longer the case.

5. Residuals are homoscedastic. This assumption states that the residuals have equal variance no matter what levels of the predictors we might choose to look at. This assumption would be violated if, for instance, the residual variance increased or decreased with the predicted value of *y*. Again, this assumption affects only inference and does not bias the regression coefficients.

6. Residuals are independent. As noted previously, this assumption requires that there is no remaining relationship between the *y* values of any two units in the sample after the predictors have been accounted for.

Potential Violations of Assumptions

- Of these, the assumption of independence is particularly dubious given the nesting of students within schools.
- Homoscedasticity may also be violated, if the effect of SES actually varies over schools.

19

Three Key Problems When Violating Independence

If the observations are positively correlated:

1. F-statistics tend to be too large.
   - We are likely to overestimate the significance of the model as a whole (the test of r-squared).

2. Standard error estimates tend to be too small.
   - We are likely to overestimate the significance of specific regression coefficients.

3. Ignoring hierarchical structure of data severely limits our ability to model within and between group effects that may be of key interest.

20

The regression model assumes that we have N unique pieces of information (minus the $p + 1$ parameter estimates). It uses this assumption to compute the F-value, the standard errors and t-values, and the degrees of freedom for the tests. If independence is violated, however, not all N cases are truly providing unique information, and the *effective sample size* is actually smaller than N. Thus our standard errors are too small and our test statistics too big. We may thereby conclude effects exist that do not.

Why Inferences Are Biased

- To get a better informal sense of why inferences are biased when the independence assumption is wrong, consider the following scenario:
  - You go to your regular physician thinking you have a cold and she diagnoses you with a serious disease. Which of these should you do?
    - Seek a second opinion from her partner in the clinic, with whom she may have consulted
    - Seek a second opinion from a physician at another clinic
- In the latter case, you have two truly independent opinions and can have greater confidence if they converge.

21

To gain a more intuitive grasp of the importance of independence, consider the scenario above. Would you rather seek a second opinion from a physician in the same practice, with whom your physician may have already consulted and discussed your case, or would you rather seek a second opinion from an independent physician? If the independent physician agrees with the original diagnosis, this ought to persuade you more than if the physician in the next office does.

1.3 Methods for Modeling Dependent Data Structures

Objectives

- Review three approaches to modeling nested data structures:
  - adjusting standard errors and test statistics
  - the fixed-effects approach
  - the multilevel approach

23

Adjusting the Results

- One approach for dealing with dependence is to correct the standard errors and test statistics.
- This approach is useful if you are interested only in overall effects (aggregated over all groups) and are not interested in evaluating how effects may vary across groups.
- Example approaches include the use of modified sandwich estimators and Huber-White estimators.

24

One method for dealing with dependent data is to simply correct the standard errors and test statistics so that they will provide valid inferences. Sometimes this correction is called a modified sandwich estimator or a White or Huber-White estimator. This estimator normally corrects for heteroscedasticity, but the modification allows for the estimation of accurate standard errors even with cluster-correlated (nested) data.

This approach is reasonable if your only concerns are with understanding the effects of the predictors aggregating over all the groups in the sample. It does not, however, allow you to disentangle within- versus between-group effects of your predictors, nor does it permit you to examine the extent to which the effects of your predictors vary over groups.

The Fixed-Effects Approach

- Another approach is to account for the dependence by adding the grouping variable as another predictor.
- For example, for the HSB data, you would include school as a classification variable in the model (requiring 160 dummy variables).

$$\text{Math}_i = \sum_{j=1}^{160} \beta_j \text{School\_Dummy}_i + \beta_{161}\text{SES}_i + r_i$$

- Each school thus has its own intercept, accounting for differences across schools in levels of achievement.
- Notice that school is assumed to be a fixed factor in the model.

25

The fixed-effects approach is popular in some areas of research, such as econometrics. The basic idea is to incorporate the source of dependence into the model so that the residuals will be uncorrelated. In the HSB example, the source of dependence is the clustering of students within schools. If schools differ in their average levels of math achievement, then this will produce residuals that are positively correlated within schools. But if these differences are controlled for by adjusting for school differences in achievement, then the residuals should be uncorrelated.

To use the fixed-effects approach, you simply generate a sequence of dummy variables, one for each group, and then enter them into the model. The general intercept of the equation is omitted so that the effect of each dummy variable represents the intercept for that particular school. In the HSB example, because there are 160 schools, we would generate 160 dummy variables, one for each school (or let SAS do this for us with a CLASS statement in the GLM procedure). Each student would be given a score of 1 on the dummy variable for their school and a score of 0 for the 159 remaining dummy variables.

SES can be included in the model as before, but note that now the effect of SES is estimated net of the school dummies. Importantly, these school dummies effectively serve as proxies for *all* characteristics that vary across schools, so we have controlled for *everything* that differs across schools in assessing the SES effect, and we need not worry about potentially important omitted variables at the school level. This is a very nice feature of this approach.

Advantages of the Fixed-Effects Approach

- Is the most appropriate method if grouping variable is truly fixed and not randomly sampled from population
- Useful when there are a small number of level-2 units
  - e.g., 50 employees nested within each of 6 stores
- Adjusts model for all unmeasured level-2 covariates
  - e.g., estimating and removing main effects of group controls for all characteristics of group
- Allows for estimation within the standard general linear modeling framework
  - e.g., PROC GLM, PROC REG, etc.

26

Depending upon the particular design and research questions, there are several advantages to the fixed-effects approach.

Probably of greatest importance, this approach allows for the explicit incorporation of information about Level 2 group membership in designs where there are a small number of Level 2 groups.

As we will see later, the number of independent groups at Level 2 has an important impact on our ability to obtain valid and reliable estimates of between group variability. If the number of groups is small (e.g., fewer than 10), we may not be able to empirically estimate variance components for the Level 2 model. In this case, the fixed-effects approach offers a practical alternative that does not encounter similar estimation problems.

Importantly, the fixed-effects approach is the most appropriate analytic method for evaluating grouping variables that are truly fixed (e.g., fixed factors in an experimental design). Here, however, we are concerned with the application of the fixed-effects approach when groups are randomly sampled from a population.

Disadvantages of the Fixed-Effects Approach

- Difficult to evaluate effects of group-level variables (can be done with contrasts, but becomes cumbersome)
- Many nuisance parameters to estimate (not parsimonious)
- Grouping variable assumed to be a fixed predictor, so inferences must be restricted to these particular groups (not a broader population of groups)

27

One problem with the fixed effects approach, however, is that we can no longer directly assess the effects of particular school characteristics, such as sector. That is, the following model would not be estimable because the sector variable is perfectly collinear with the set of school dummies:

$$\text{Math}_i = \sum_{j=1}^{160} \beta_j \text{School\_Dummy}_i + \beta_{161}\text{SES}_i + \beta_{162}\text{Sector}_i + r_i$$

Of course, given this collinearity, we could use contrasts with the school dummy coefficients to compare the average intercepts of private and public schools, but this would be quite tedious. Further, using contrasts would be cumbersome with many or continuous school-level predictors.

There are other important drawbacks to the fixed-effects approach as well. For one thing, it requires that we estimate 160 intercepts, so it is not especially parsimonious. In this case we have over 7000 students, so we can accommodate all these extra parameters, but if only a few students were sampled per school this would present a greater problem. In addition, if we believe that the SES effect may vary across schools, we would have to interact SES with the 160 dummies, resulting in 320 coefficients to be estimated!

Another drawback is that the fixed-effects approach (true to its name) assumes the grouping variable is a fixed predictor, just like any other regression model. This implies that specific schools have been selected for study, and we are interested in making inferences about students in these specific schools. However, often it is more reasonable to think of an experimental design where schools are randomly sampled from a population, making this a random rather than fixed factor. Treating school as random will also allow us to make inferences to the entire population of schools from whence our specific 160 schools were drawn. In other words, our inferences are not limited to the specific schools in our sample.

To allow for random effects of schools, we must move out of the standard regression model and into the multilevel regression model.

Multilevel Modeling

- Multilevel modeling does not incorporate schools as a fixed effects predictor, but rather treats schools as randomly sampled from a population.
- Effects are not estimated individually for each school, but are assumed to have a particular distribution across the *population* of schools.
- We earlier estimated the OLS regression model:

$$\text{Math}_i = \beta_0 + \beta_1 \text{SES}_i + r_i$$

- We can write the regression model as

$$\text{Math}_{ij} = \beta_{0j} + \beta_1 \text{SES}_{ij} + r_{ij}$$

where we will assume a particular distribution for β_{0j} just as we customarily do for r_{ij} .

28

In the multilevel modeling approach, we assume that there are two levels of sampling, the individual (micro-unit) and the school (macro-unit). As such, we now need two subscripts to track the nesting of individuals (i) within schools (j). Note that by placing the j subscript on the intercept parameter, β_{0j} , we are indicating that the intercept varies across schools, similar to the fixed-effects approach.

Unlike the fixed-effects approach, in the multilevel approach the specific schools in the HSB are treated as a random sample from all possible schools. Hence, the intercepts for the schools in our sample are also randomly drawn from the distribution of all school intercepts. This is the origin of the term *random effect* – it is an effect that varies randomly (according to some distribution) in the population. Rather than try to estimate an intercept for every school in our sample (as in the fixed-effects approach) we will instead attempt to estimate the *distribution* of the random intercept in the population using our sample data.

For instance, we might assume that, in the population of high schools, the intercepts are normally distributed. We would then try to estimate the mean and variance of the β_{0j} distribution. The assumption of normality for random effects is quite standard, though other distributions have been entertained in the literature (for example, semiparametric distributions or mixtures of normals).

Advantages of Multilevel Models

- Parsimonious
- Can make inferences to the population of groups
- Conform to the sampling design with the random selection of groups followed by the random selection of individuals within groups
- Enable us to examine the effects of individual-level and group-level influences simultaneously

29

There are several advantages of the multilevel, or random-effects, approach relative to the fixed-effects approach. One is parsimony. The fixed-effects approach requires the estimation of 160 unique school intercepts, while the random-effects approach requires only the estimation of a mean and variance for the distribution of school intercepts. (Individual school estimates can, however, be obtained based on the model estimates). Of course, this requires that the normality assumption for the random effects is reasonable.

A second advantage is that multilevel models allow us to make inferences to the population of all schools, as opposed to restricting our inference to students within the 160 schools that happened to be sampled in the HSB. Because we typically want to make inferences to this broader population, this is a key advantage of multilevel models.

Third, multilevel models often conform more closely to the substantive theory of interest. In the fixed-effects approach, the school intercepts are nuisance parameters, whereas in the multilevel approach these differences are of key interest. Theoretically, the idea of nesting within multilevel models is a powerful one that can help to guide our thinking about dependence and random effects.

The last difference between the two approaches is perhaps the most important. Recall that including the school dummies in the fixed-effects approach also served to control for all systematic differences between schools. The advantage of this was that we no longer had to worry about potentially important but omitted school-level predictors. The disadvantage, however, was that we could no longer easily assess the importance of particular school-level predictors (except in some cases using contrasts). Multilevel models differ on both of these points. The omission of potentially important school-level variables can be problematic, but we can directly and easily assess the importance of specific school-level predictors such as Sector or School Size. When we are interested in assessing context effects, this aspect of multilevel models is a key advantage.

1.4 The Random-Effects ANOVA Model

Objectives

- Understand the random-effects ANOVA model.
- Introduce the MIXED procedure.
- Estimate the random-effects ANOVA model in PROC MIXED.
- Calculate and interpret the intraclass correlation coefficient (ICC).

31

The Random-Effects ANOVA Model

- The simplest multilevel model is a random-effects ANOVA.
- There are no predictors in the model, only a random intercept.
- The random intercept captures mean differences between the groups.
- Like the fixed-effects ANOVA, we want to decompose the variance of the observed variable into within- and between-group components.

32

Recall that in a standard one-way ANOVA model we include a grouping variable as a single fixed factor with J levels. We then evaluate the portion of variance in the outcome that is due to differences of the group means from the grand mean (between-groups variance) relative to the portion that is due to differences of the individuals' scores from their respective group means.

In random-effects ANOVA, we do much the same thing, except that the grouping variable is not fixed but rather is treated as a level of nesting for the data. The intercept parameter is then also treated as random, reflecting group-mean differences (the intercept is the mean in the absence of other predictors). We will estimate a variance for the group means and use this to decompose the variance of the dependent variable into its within- and between-group components.

The Random-Effects ANOVA Model

Level 1: $y_{ij} = \beta_{0j} + r_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + u_{0j}$

Reduced Form: $y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$

33

Because the data has been randomly sampled at two levels, we will write a model for each level. Level 1 is the lowest level of the model, typically corresponding to the individual. Level 2 is the highest level of the model, typically corresponding to groups.

For the Level 1 model we have:

y_{ij} The observed value of the dependent variable for individual *i* belonging to group *j*.

β_{0j} The random intercept parameter. Because there are no predictors at Level 1, the random intercepts correspond to the group means.

r_{ij} The residual for individual *i* within group *j*. In the random-effects ANOVA model, this represents the difference between y_{ij} and β_{0j}, or variability about the group mean.

At Level 2, we treat the random effects at Level 1 as if they are outcome variables. Here, we express β_{0j} as a function of two values:

γ_{00} The intercept of the β_{0j} equation. Because there are no predictors, this simply represents the overall mean of the group means, or the group mean for an average group (where u_{0j} is zero).

u_{0j} The residual for the β_{0j} equation. Because there are no predictors, this simply represents the difference between β_{0j} and γ_{00}.

The reduced-form equation shows the decomposition of *y* into three parts: the overall mean of the group means, the difference between the group mean and the overall mean, and the difference between the individual's score and their group mean.

The Variance Components of the Model

- The reduced form equation is

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

- It includes one fixed effect, the grand mean, and two random components, the residual at Level 1 and the residual at Level 2.
- We will assume that these residuals are normally distributed and uncorrelated with one another:

$$r_{ij} \sim N(0, \sigma^2) \qquad u_{0j} \sim N(0, \tau_{00})$$

- This implies that

$$E(y_{ij}) = (\gamma_{00})$$
$$V(y_{ij}) = V\left(u_{0j} + r_{ij}\right) = V(u_{0j}) + V(r_{ij}) = \tau_{00} + \sigma^2$$

34

In general, the reduced-form equation of a multilevel model is composed of fixed effects (the γ's) and random effects (u and r).

We must make assumptions about how the random effects are distributed. Typical assumptions for linear multilevel models are these:

1. The residuals at both levels of the model are normally distributed.
2. The residuals at Level 2 are uncorrelated with the residuals at Level 1.

More formally, for the random-effects ANOVA we can write the first of these assumptions as

$$r_{ij} \sim N(0, \sigma^2)$$
$$u_{0j} \sim N(0, \tau_{00})$$

σ^2 The variance of the Level 1 residuals, r_{ij}.

τ_{00} The variance of the Level 2 residuals, u_{0j}.

✎ The notation $r_{ij} \sim N(0, \sigma^2)$ reads "r_{ij} is normally distributed with mean zero and variance σ^2."

The first of these statements is quite familiar to us, as it is the same assumption that we make in the standard one-way ANOVA model – we assume that the scores within each group are normally distributed with the same variance for every group. The second assumption is new and arises from the treatment of groups as a random factor. This assumption states that, in the population, the group means are normally distributed around the grand mean with variance τ_{00}. We want to estimate the variance in the population using our particular sample of groups.

The Intraclass Correlation

- Because the total variance is decomposed into two additive components, we can calculate the portion due to between-group mean differences as

$$\text{ICC} = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

- This value is referred to as the intraclass correlation because it also represents the degree of correlation between individuals within a group (or class).
- The intraclass correlation measures the degree of dependence in the data, or the strength of the nesting effect.
- Note that the general linear model assumes ICC=0.

35

The most important statistic that can be estimated from the random-effects ANOVA model is the intraclass correlation. This is similar to the R^2 calculated in a standard one-way ANOVA model. We simply take the between-groups variance and divide by the total variance (the sum of the between-groups and within-groups variance).

It thus measures the proportion of variance in *y* due exclusively to differences between the groups. The remainder (1-ICC) is the proportion of variance that exists between individuals within groups.

The reason this statistic is referred to as the intraclass correlation is that it can also be interpreted as the correlation between the *y* values of any two individuals who share a group.

It is thus a direct measure of the degree of dependence in the data and hence the strength of the effect of the nesting structure.

Estimating Multilevel Models in SAS

- Linear multilevel models, such as the random-effects ANOVA model, can be estimated in SAS using the MIXED procedure.
- The reduced form equation is used to define the model in MIXED.
- It is often more difficult to begin from the reduced form equation compared to the Level 1 and 2 expressions.
- A good general strategy for specifying multilevel models in MIXED is to first write out the Level 1 and Level 2 equations and then construct the reduced form equation from these expressions.

36

We strongly advise that before specifying the model in SAS you begin by writing the Level 1 and Level 2 equations and then combine these to create the reduced-form equation. In our experience, many errors in model specification and interpretation occur when analysts attempt to write the model in reduced-form directly (and specify the model in SAS) without first writing out the full set of model equations.

The MIXED Procedure

PROC MIXED *options*;
    CLASS *classification variables;*
    MODEL *outcome = fixed-effect predictors / options*;
    RANDOM *random effects / options*;
    ESTIMATE *'label' fixed-effects values* **|**
        random-effects values / options;
    CONTRAST *'label' fixed-effects values* **|**
        random-effects values / options;
RUN;

37

Selected MIXED procedure statements:

CLASS — names the classification variables to be used in the analysis. If the CLASS statement is used, it must appear before the MODEL statement. Classification variables can be either character or numeric.

MODEL — is a required statement that names a single dependent variable (to the left of the equal sign) and the predictors with fixed effects (to the right of the equal sign). A fixed intercept parameter is included by default.

RANDOM — defines the random effects. Variables with random effects can be classification or continuous variables, and multiple RANDOM statements are possible.

ESTIMATE — estimates linear combinations of fixed or random effects.

CONTRAST — tests composite hypotheses concerning linear combinations of fixed and/or random effects.

✎ Including the Level 2 ID variable in the CLASS statement is required if the data is either (1) not sorted by Level 2 ID or (2) if the Level 2 ID is character variable (e.g., GROUP01, GROUP02, etc.). However, if the Level 2 ID is both numeric and sorted prior to the analysis, the CLASS statement is not required and in turn increases programming speed.

Fitting a Random-Effects ANOVA with PROC MIXED

c1_hsb_reanova.sas

For this demonstration we will continue our analysis of the HSB data described in Chapter 1. Recall that we estimated a multiple regression model that ignored potential nesting in the data. We now want to calculate the intraclass correlation between students within a school to determine the degree of dependence present in the data. Prior to calculating the ICC, it is often helpful to graphically examine the distribution of the sample group means.

Exploring Group-Specific Means

We will begin by looking at the mean of student math achievement within each of the 160 schools.

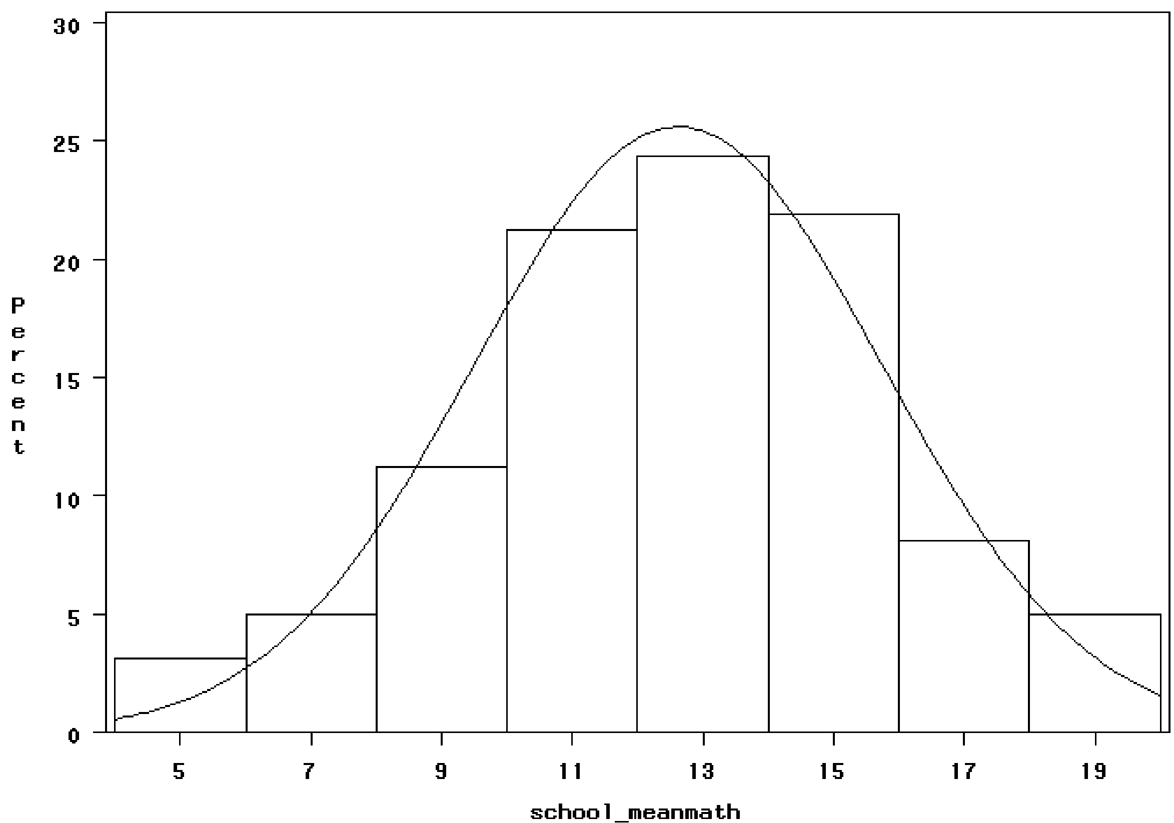
```
proc means data=mixed.hsb mean std min max;
   var student_mathach;
   class school_id;
run;
```

Analysis Variable : student_mathach

school_ID	N Obs	Mean	Std Dev	Minimum	Maximum
1224	47	9.7154468	7.5927847	-2.8320000	23.5840000
1288	25	13.5108000	7.0218429	1.5750000	23.5780000
1296	48	7.6359583	5.3510703	-1.3530000	23.1720000
<snip>					
2467	52	10.1475192	6.7828033	-2.4220000	24.0120000
2526	57	17.0530000	4.7760522	2.8100000	24.9930000
2626	38	13.3966053	6.2426489	-2.5440000	24.9930000
<snip>					
9508	35	13.5746571	6.4640023	0.2940000	23.5010000
9550	29	11.0891379	7.8779977	-2.2310000	23.8180000
9586	59	14.8636949	6.4159995	-2.8320000	24.9930000

It appears that there may be some variability of means across the schools. We can also plot these means.

```
proc means data=mixed.hsb noprint;
   var student_mathach;
   by school_id;
   output out=mathmeans mean=school_meanmath;
run;
goptions reset=all hsize=7 vsize=5;
proc univariate data=mathmeans;
   var school_meanmath;
   histogram/normal(color=black);
run;
```



There appears to be variability in mean math achievement across schools. Recall that the grand-mean math achievement is approximately 12.7. The histogram shows that, as expected, the most of the school means are massed around the grand mean. However, there are also schools with both substantially lower and substantially higher mean math achievement scores.

The temptation is to conclude that there is meaningful variability across the group means based on this distribution. However, we do not yet know how the variability across schools compares to the variability within schools. These observed differences might be nothing more than random sampling variability. To more formally evaluate this, we will turn to the random-effects ANOVA model and the ICC.

Random-Effects ANOVA

We will always define the model in PROC MIXED in reduced form. For the random-effects ANOVA model we have:

Level 1 Equation:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij} \qquad r_{ij} \sim N(0, \sigma^2)$$

Level 2 Equation:

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad u_{0j} \sim N(0, \tau_{00})$$

Reduced-Form Equation:

$$\text{Math}_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

We see that we have one fixed effect, the fixed intercept (γ_{00}), plus the random component to the intercept (u_{0j}), plus the Level 1 residual r_{ij}.

Thus, the three parameters that we need to estimate are γ_{00}, τ_{00}, and σ^2.

We begin by sorting the data by **`school_id`** so that the data is ordered properly for the MIXED analyses. Note that we do not need to manually sort the data if we use a CLASS statement in PROC MIXED. However, each time MIXED encounters the CLASS statement, the data is re-sorted; if the data is already sorted, omitting the CLASS statement can increase computational efficiency.

```
proc sort data=mixed.hsb;
   by school_id;
run;
```

We will specify a random ANOVA model using the following code:

```
proc mixed data=mixed.hsb method=reml cl covtest;
   model student_mathach = / solution alpha=.05 ddfm=bw;
   random intercept/subject=school_id v vcorr;
run;
```

The options selected for the MIXED statement are **method=reml**, **cl**, and **covtest**. The first of these specified that restricted maximum likelihood is to be used to estimate the model. This is the recommended estimator for multilevel models. The second requests confidence limits for the variance component estimates, and the third requests significance tests of the variance component estimates.

The MODEL statement specifies the fixed effects for the model. In this case, there are no predictors with fixed effects, just the default fixed intercept. Here are selected MODEL statement options:

SOLUTION — requests that the estimates, standard errors, t-statistics, degrees of freedom, and p-values be displayed for all fixed-effects.

ALPHA=.05 — requests that 95% t-type confidence intervals be produced for each fixed-effect estimate.

DDFM=BW — requests that the degrees of freedom for testing the fixed effects be computed using the between and within method. This method is what is typically used for multilevel models, but better methods are available for small samples, like Kenward-Rogers (DDFM=KR).

RANDOM — specifies that the intercept is to be a random effect.

SUBJECT — This option is necessary to indicate the nesting structure of the data. Defining **SUBJECT=school_id** tells PROC MIXED that the intercept is to vary randomly across schools.

V — requests that the covariance matrix among Level 1 residuals be printed.

VCORR — requests that the correlation matrix among Level 1 residuals be printed (this is the standardized V matrix). For the random-effects ANOVA model, the off-diagonal element will be the intraclass correlation.

PROC MIXED Output:

```
                            The Mixed Procedure

                            Model Information

          Data Set                      MIXED.HSB
          Dependent Variable            student_mathach
          Covariance Structure          Variance Components
          Subject Effect                school_ID
          Estimation Method             REML
          Residual Variance Method      Profile
          Fixed Effects SE Method       Model-Based
          Degrees of Freedom Method     Between-Within
```

The Model Information section is useful for making sure that you have specified the model as intended. Variance Components is the default structure for covariance matrix of the random effects at Level 2. It specifies that the random effects are independent. In this case, this assumption is fine because we have one random effect (the intercept).

```
                              Dimensions

                  Covariance Parameters             2
                  Columns in X                      1
                  Columns in Z Per Subject          1
                  Subjects                        160
                  Max Obs Per Subject              67


                        Number of Observations

              Number of Observations Read           7185
              Number of Observations Used           7185
              Number of Observations Not Used          0
```

The Dimensions section provides information on the size of the model and data set. The two covariance parameters correspond to τ_{00} and σ^2. In PROC MIXED, the X matrix is the design matrix for the fixed effects. It will have as many columns as there are fixed effects in the model. Because our model includes only one fixed effect, γ_{00}, this is listed as one. The Z matrix is the design matrix for the random effects at Level 2. We have a random intercept, so this is also listed as one. Last, PROC MIXED is reporting that there are 160 subjects (independent sampling units). This is the number of unique values for **`school_id`** in the data. The maximum number of observations per subject (students per school) is 67.

```
                           Iteration History

        Iteration    Evaluations    -2 Res Log Like       Criterion

                0              1     48102.91726234
                1              2     47116.81230623      0.00000109
                2              1     47116.79350024      0.00000000

                      Convergence criteria met.
```

The Iteration History section describes the optimization of the model. The message “Convergence criteria met” indicates that the model has converged. *If the model does not converge, do not interpret the model estimates*. You may need to increase the number of iterations that PROC MIXED will perform, use different start values (input by the PARMS statement), or there may be a problem with your model.

By default, the V and VCORR option provides the estimate of the intraclass covariance and correlation matrix for the first group, respectively. The V and VCORR matrices are the same across all groups, so we only need to consider one. Here are the first seven rows and columns:

Estimated V Matrix for Subject 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	47.7584	8.6097	8.6097	8.6097	8.6097	8.6097	8.6097
2	8.6097	47.7584	8.6097	8.6097	8.6097	8.6097	8.6097
3	8.6097	8.6097	47.7584	8.6097	8.6097	8.6097	8.6097
4	8.6097	8.6097	8.6097	47.7584	8.6097	8.6097	8.6097
5	8.6097	8.6097	8.6097	8.6097	47.7584	8.6097	8.6097
6	8.6097	8.6097	8.6097	8.6097	8.6097	47.7584	8.6097
7	8.6097	8.6097	8.6097	8.6097	8.6097	8.6097	47.7584

Estimated V Correlation Matrix for Subject 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	1.0000	0.1803	0.1803	0.1803	0.1803	0.1803	0.1803
2	0.1803	1.0000	0.1803	0.1803	0.1803	0.1803	0.1803
3	0.1803	0.1803	1.0000	0.1803	0.1803	0.1803	0.1803
4	0.1803	0.1803	0.1803	1.0000	0.1803	0.1803	0.1803
5	0.1803	0.1803	0.1803	0.1803	1.0000	0.1803	0.1803
6	0.1803	0.1803	0.1803	0.1803	0.1803	1.0000	0.1803
7	0.1803	0.1803	0.1803	0.1803	0.1803	0.1803	1.0000

This reflects that the estimated ICC is equal to .1803. Note that not only is this correlation equal across all individuals within the first group, but it is also equal across all groups (only the groups are assumed to be independent of one another). This single within-class correlation highlights the Level 1 correlation structure that is imposed by the random intercept model (e.g., compound symmetry).

Our Level 1 and Level 2 variance estimates and test statistics are shown here:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z	Alpha
Intercept	school_ID	8.6097	1.0778	7.99	<.0001	0.05
Residual		39.1487	0.6607	59.26	<.0001	0.05

Cov Parm	Subject	Lower	Upper
Intercept	school_ID	6.8339	11.1843
Residual		37.8855	40.4765

The Covariance Parameter Estimates section provides estimates of our variance components, which are

$\hat{\tau}_{00} = 8.61$ and $\hat{\sigma}^2 = 39.15$.

Thus, our estimate of the ICC is $\dfrac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \hat{\sigma}^2} = \dfrac{8.61}{8.61 + 39.15} = .18$

18% of the variance in achievement scores is estimated to be due to between-school differences and 82% is due to differences among students within schools. Put another way, the correlation between the achievement scores of students attending the same school is .18. This is of course the same value as was presented in the VCORR matrix.

In addition, we requested confidence limits and null hypothesis tests for the covariance parameter estimates by including CL and COVTEST in the PROC MIXED statement. For variances, the confidence limits take into account the lower boundary of zero and are computed based on a χ^2 distribution. From these intervals, we can see that our precision is greater for the variance component at Level 1 than the variance component at Level 2, a result that is quite typical for multilevel models. These confidence intervals will never include zero, so in practice very small values for the lower bound may indicate that the variance component is unnecessary. Alternatively, one can use the tests supplied by COVTEST to assess the variance components; in this case, both are statistically significant.

✎ For covariances and other parameters without a lower boundary, the confidence limits computed using the CL option are based on a normal distribution.

✎ COVTEST produces Wald z-tests and p-values for all variance and covariance parameter estimates; however, these tests assume asymptotic (large-sample) normality of the estimates. Given the lower boundary of zero for variance parameters, the sampling distributions for these parameter estimates tend to be skewed unless samples are extremely large, making these tests inaccurate.

✎ It is important to recognize that because the p-values and confidence limits are based on alternative assumptions about the sampling distributions of the variance components, they will not always agree (that is, the χ^2 based confidence limits may exclude zero, yet a normal-theory null hypothesis test *not* be rejected by the p-value).

✎ There is no direct test of the ICC, but note that the ICC is zero when τ_{00} is zero, so typically the test of $\hat{\tau}_{00}$ is used as a proxy for a test of the estimated ICC.

```
                      Solution for Fixed Effects

                           Standard
Effect           Estimate      Error     DF    t Value   Pr > |t|    Alpha

Intercept         12.6370     0.2443    159     51.72     <.0001     0.05

                      Solution for Fixed Effects

                 Effect               Lower       Upper

                 Intercept          12.1544     13.1195
```

We see that our fixed intercept estimate, which is our estimate of the average group mean, is $\hat{\gamma}_{00} = 12.64$ with a 95% t-type confidence interval spanning from 12.15 to 13.12. This value differs slightly from the grand mean. Note that if the groups differ in size, as in this case, the average of the group means will not necessarily be identical to the grand mean.

This can be seen by comparing the intercept estimate above to the grand mean:

```
proc means data=mixed.hsb n mean var min max;
   var student_mathach;
run;
```

```
                                                     The MEANS Procedure

                  Analysis Variable : student_mathach

   N            Mean        Variance         Minimum         Maximum
 ---------------------------------------------------------------------
7185      12.7478526      47.3102637      -2.8320000      24.9930000
 ---------------------------------------------------------------------
```

As in this case, the difference is usually quite small.

1.5 Random-Effects Regression

Objectives

- Incorporate predictors at Level 1.
- Understand the implications of allowing Level 1 predictors to be fixed versus random.
- Demonstrate the fitting of a random-effects regression model in PROC MIXED.

3

In this section of the course we examine the differences between three multilevel models incorporating a single predictor at Level 1. To highlight these differences we will do the following:

- Write the Level 1 and Level 2 equations for the model.
- Graphically consider the implications of choosing different parameters of the regression model to be random.
- Examine the reduced-form equation for the model that we will use to specify the model in PROC MIXED.
- Provide generic code for the specification of these models in PROC MIXED.

One Predictor: Fixed Intercept and Fixed Slope

Level 1 Model:

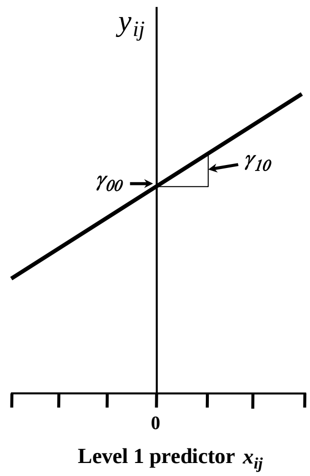
$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij}$$

Level 2 Model:

$$\beta_{0j} = \gamma_{00}$$
$$\beta_{1j} = \gamma_{10}$$

Reduced-Form Model:

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + r_{ij}$$



4

One Predictor: Fixed Intercept and Fixed Slope

Level 1 Model:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij}$$

Level 2 Model:

$$\beta_{0j} = \gamma_{00}$$
$$\beta_{1j} = \gamma_{10}$$

Reduced-Form Model:

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + r_{ij}$$

$$r_{ij} \sim N(0, \sigma^2)$$

```
proc mixed data=mydata method=reml cl covtest;
     model y = x / solution alpha=.05;
run;
```

5

This model is equivalent to the simple regression model (estimated by residual/restricted maximum likelihood) where each individual is assumed to be an independent sampling unit.

One Predictor: Random Intercept and Fixed Slope

Level 1 Model:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + r_{ij}$$

Level 2 Model:

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

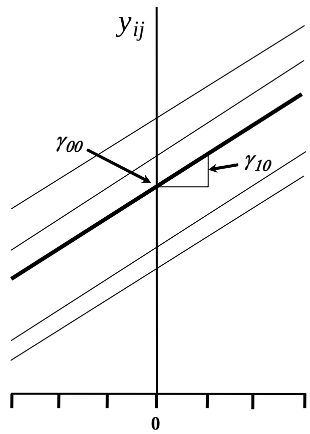
$$\beta_{1j} = \gamma_{10}$$

Reduced-Form Model:

$$y_{ij} = \left(\gamma_{00} + u_{0j}\right) + \gamma_{10}x_{ij} + r_{ij}$$

$$= \left(\gamma_{00} + \gamma_{10}x_{ij}\right) + u_{0j} + r_{ij}$$

Random Intercept, Fixed Slope



Level 1 predictor x_{ij}

6

One Predictor: Random Intercept and Fixed Slope

Level 1 Model:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + r_{ij}$$

Level 2 Model:

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

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

Reduced-Form Model:

$$y_{ij} = \left(\gamma_{00} + u_{0j}\right) + \gamma_{10}x_{ij} + r_{ij}$$

$$= \left(\gamma_{00} + \gamma_{10}x_{ij}\right) + u_{0j} + r_{ij}$$

$$r_{ij} \sim N(0, \sigma^2)$$

$$u_{0j} \sim N(0, \tau_{00})$$

```
proc mixed data=mydata method=reml cl covtest;
     model y = x / solution alpha=.05;
     random intercept / subject=L2_ID;
run;
```

7

Note that we now have two variance components in the model, the random intercept and the Level 1 residual. The SUBJECT option in the RANDOM statement indicates that our independent sampling units (Level 2 units) are indicated in the data by the **L2_ID** variable.

One Predictor: Random Intercept and Random Slope

Level 1 Model:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + r_{ij}$$

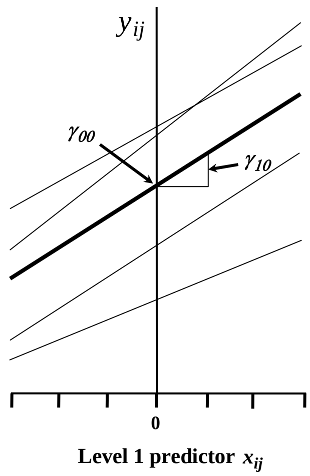
Level 2 Model:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Reduced-Form Model:

$$y_{ij} = (\gamma_{00} + u_{0j}) + (\gamma_{10} + u_{1j})x_{ij} + r_{ij}$$
$$= (\gamma_{00} + \gamma_{10}x_{ij}) + (u_{0j} + u_{1j}x_{ij}) + r_{ij}$$

Random Intercept, Random Slope



8

Covariance Matrix of Random Effects: Tau

$$\text{var}(r_{ij}) = \sigma^2$$ Level 1

$$\text{var}(u_{0j}) = \tau_{00}$$
$$\text{var}(u_{1j}) = \tau_{11}$$
$$\text{cov}(u_{0j}, u_{1j}) = \tau_{10} = \tau_{01}$$ Level 2

$$\mathbf{T} = \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}$$ Covariance matrix

9

Note that we now have two random effects, a random intercept and a random slope, so we must consider not just the variances of these random effects but also their covariance. We will designate the covariance matrix of the random effects as **T** (Tau). Here, **T** is a 2 × 2 symmetric matrix, with three unique parameters to be estimated.

One Predictor: Random Intercept and Random Slope

Level 1 Model:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij}$$

Level 2 Model:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

Reduced-Form Model:

$$y_{ij} = (\gamma_{00} + u_{0j}) + (\gamma_{10} + u_{1j}) x_{ij} + r_{ij}$$
$$= (\gamma_{00} + \gamma_{10} x_{ij}) + (u_{0j} + u_{1j} x_{ij}) + r_{ij}$$

$$r_{ij} \sim N(0, \sigma^2)$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$$

```
proc mixed data=mydata method=reml cl covtest;
   model y = x / solution alpha=.05;
   random intercept x / subject=L2_ID type=un gcorr;
run;
```

10

Note that the RANDOM statement now includes both **`intercept`** (the random intercept) and **`x`** (the random slope). The TYPE option overrides the default assumption of independence of these random effects. For many mixed models, an assumption of independence for the random effects is realistic, but for multilevel models it typically is not. As such, we have told SAS to change this matrix from a diagonal (independence) matrix to an unstructured matrix. It will now estimate all three unique elements in the **T** matrix, the variance of the intercept, variance of the slope of *x*, and covariance between the two.

The GCORR option tells SAS to rescale the covariance matrix of the random effects into a correlation matrix. The associations among the random effects are typically easier to interpret as correlations than as covariances.

✎ In PROC MIXED, the **T** matrix (covariance matrix of the random effects) is referred to as the **G** matrix (hence the option name GCORR).

✎ The type UN option for the random effects, while intuitive, is not the only specification available. For instance, TYPE=FAO(2) (for two random effects) can be used to constrain **T** to be nonnegative definite.

Fitting a Random-Effects Regression with PROC MIXED

c2_hsb_rereg.sas

For this demonstration we will again be considering the HSB data set first introduced in Chapter 1, **mixed.hsb**. Here we will fit the three regression models described above and compare their substantive implications. For these analyses, our dependent variable will again be math achievement, **student_mathach**, and our predictor will be the socioeconomic status of the student's family, **student_ses**. The nesting of students within schools is indicated by **school_ID**.

Simple Regression Model

We will begin with the simple regression model with a fixed intercept and slope:

Level 1 Equation:

$$\text{Math}_{ij} = \beta_{0j} + \beta_{1j}\text{SES}_{ij} + r_{ij} \qquad r_{ij} \sim N(0, \sigma^2)$$

Level 2 Equations:

$$\beta_{0j} = \gamma_{00}$$
$$\beta_{1j} = \gamma_{10}$$

Reduced-Form Equation:

$$\text{Math}_{ij} = \underbrace{\gamma_{00} + \gamma_{10}\text{SES}_{ij}}_{\text{Fixed-Effects}} + r_{ij}$$

To enter this model into the MIXED procedure, we write:

```
proc mixed data=mixed.hsb method=reml cl covtest;
   model student_mathach = student_ses / solution alpha=.05;
run;
```

This is, in effect, the same model that we estimated in Chapter 1 by ordinary least-squares. The only difference is that we are using restricted maximum likelihood to estimate the model. The results from this model will, however, provide a useful baseline for judging subsequent models.

PROC MIXED Output:

```
                            Model Information

           Data Set                      MIXED.HSB
           Dependent Variable            student_mathach
           Covariance Structure          Diagonal
           Estimation Method             REML
           Residual Variance Method      Profile
           Fixed Effects SE Method       Model-Based
           Degrees of Freedom Method     Residual
```

Notice that the covariance structure is diagonal – this indicates that we are assuming independence of observations.

Dimensions	
Covariance Parameters	1
Columns in X	2
Columns in Z	0
Subjects	1
Max Obs Per Subject	7185

There is one covariance parameter, σ^2. Recall that X is the design matrix of the fixed effects, including in this case the intercept and the effect of **`student_ses`**. Z is the design matrix for the random effects, which in this case is empty. Due to the lack of random effects in this model, all 7185 observations are assumed to be independent.

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Alpha	Lower	Upper
Residual	41.1588	0.6868	59.93	<.0001	0.05	39.8452	42.5388

Our estimate of the Level 1 residual variance is $\hat{\sigma}^2 = 41.16$. This is almost identical to the mean-squared error estimate from our ordinary least-squares regression model.

Solution for Fixed Effects

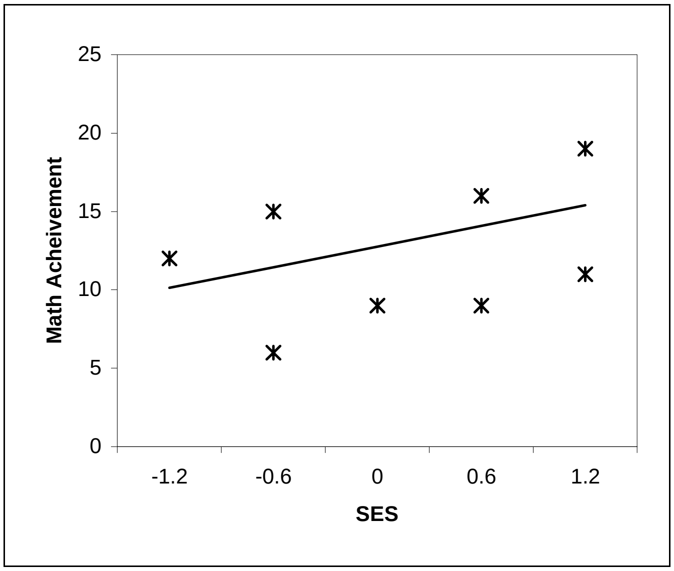
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
Intercept	12.7474	0.07569	7183	168.42	<.0001	0.05
student_ses	3.1839	0.09712	7183	32.78	<.0001	0.05

The fixed-effects estimates are nearly identical to the estimates we obtained earlier using ordinary least-squares. Because **`student_ses`** has been mean-centered (has a mean of zero), we can interpret these estimates as follows: The estimate $\hat{\gamma}_{00} = 12.75$ is the expected math achievement of a student from an average SES family; and the estimate $\hat{\gamma}_{10} = 3.18$ indicates the expected increase in math achievement per one unit increase on our SES index.

One important feature of the model that we have specified is that it assumes independence of observations. Further, it allows for no random effects – that is, there is only one intercept and one slope estimate and these estimates apply equally to all students no matter what school they may come from.

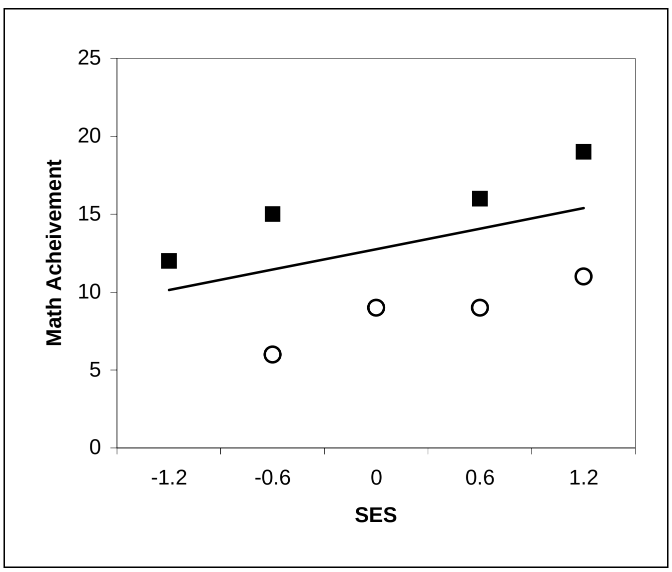
Simple Regression Model

To better understand how these two aspects of the model are related, consider a plot of eight hypothetical cases against the estimated regression line:



Correlated Residuals

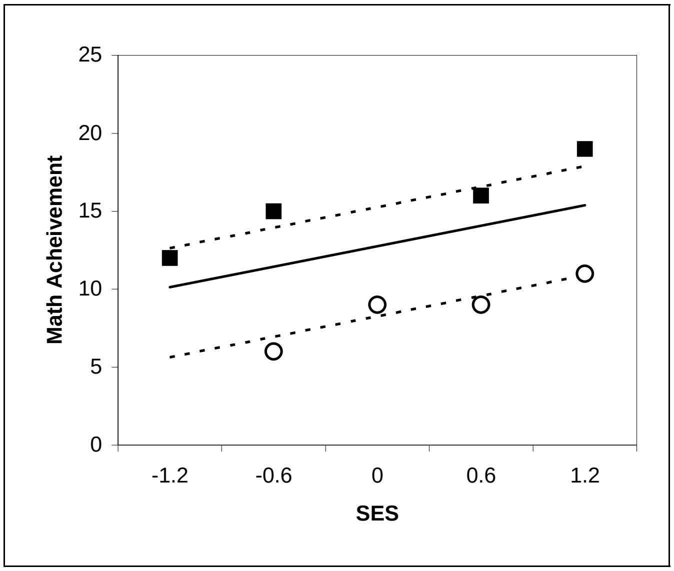
Our model assumes that these residuals are all independently distributed around the regression line with constant variance. But suppose that these cases actually come from two different schools:



Here we see that students in one school are, on average, scoring higher than students at another school. That is, there are achievement differences between schools. The residuals for one school are all positive and the residuals for the other are all negative – the residuals are thus positively correlated within schools. Indeed, from our earlier random-effects ANOVA on the same data, we know that there is a substantial variance component associated with school achievement differences, and we would like to account for this source of variability in these models as well.

Random Intercepts

To allow for school differences in levels of achievement, we can introduce a random intercept parameter, much as we did in the random-effects ANOVA. Conceptually, this means that each school will have their own intercept and the variation in these intercepts will accommodate the lack of independence in the data, as shown below.



Random Intercepts Model

Our mixed model will assume that these school-specific intercepts are normally distributed around the *average* regression line. Further, we will now assume that the residuals for the students are independently and normally distributed around their *group* regression line. These seem like much more reasonable assumptions. The model equations are shown here:

Level 1 Equation:

$$\text{Math}_{ij} = \beta_{0j} + \beta_{1j}\text{SES}_{ij} + r_{ij} \qquad r_{ij} \sim N(0, \sigma^2)$$

Level 2 Equations:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} \end{aligned} \qquad u_{0j} \sim N(0, \tau_{00})$$

Reduced-Form Equation:

$$\begin{aligned} \text{Math}_{ij} &= \left(\gamma_{00} + u_{0j}\right) + \gamma_{10}\text{SES}_{ij} + r_{ij} \\ &= \underbrace{\left(\gamma_{00} + \gamma_{10}\text{SES}_{ij}\right)}_{\text{Fixed-Effects}} + \underbrace{u_{0j}}_{\text{Random-Effect}} + r_{ij} \end{aligned}$$

To enter this model into PROC MIXED, we write the following:

```
proc mixed data=mixed.hsb method=reml cl covtest;
   model student_mathach = student_ses / solution alpha=.05 ddfm=bw;
   random intercept / subject=school_id;
run;
```

PROC MIXED Output:

```
                          Model Information

          Data Set                      MIXED.HSB
          Dependent Variable            student_mathach
          Covariance Structure          Variance Components
          Subject Effect                school_ID
          Estimation Method             REML
          Residual Variance Method      Profile
          Fixed Effects SE Method       Model-Based
          Degrees of Freedom Method     Between-Within


                              Dimensions

                    Covariance Parameters              2
                    Columns in X                       2
                    Columns in Z Per Subject           1
                    Subjects                         160
                    Max Obs Per Subject               67
```

Note that the covariance structure has been changed to a variance components specification and that we now have two covariance parameters, σ^2 and τ_{00}. We still have two fixed-effects predictors, so the columns in X are unchanged, but the number of columns in Z is now one, reflecting the random intercept.

The subjects line now lists 160 schools as the independent sampling units.

```
                       Covariance Parameter Estimates

                                           Standard         Z
Cov Parm      Subject      Estimate       Error     Value        Pr Z    Alpha

Intercept     school_ID      4.7665      0.6549      7.28      <.0001     0.05
Residual                    37.0346      0.6254     59.22      <.0001     0.05

              Cov Parm      Subject         Lower       Upper

              Intercept     school_ID      3.7045      6.3636
              Residual                    35.8388     38.2916
```

Our estimates for the variance components are $\hat{\tau}_{00} = 4.77$ and $\hat{\sigma}^2 = 37.03$ and the lower bounds of both confidence intervals are quite far from zero. Of note, the addition of the Level 2 variance component has resulted in a reduction in the Level 1 residual variance (from 41 to 37). This is to be expected, as allowing the school intercepts to vary necessarily decreases the distance between the students' observations and the regression lines for their schools.

It is also interesting to compare these estimates to the Level 1 and Level 2 variance components from the random-effects ANOVA that we estimated in the last demonstration. In the random-effects ANOVA, $\hat{\tau}_{00} = 8.61$ and $\hat{\sigma}^2 = 39.15$. Recall that the random-effects ANOVA includes no Level 1 predictors, so these are *unconditional* variances, whereas in the random-effects regression they are *conditional* on SES. Intuitively, one would expect that adding a predictor at Level 1 would decrease the variance at that level and, indeed, this variance decreases from 39 to 37 through the inclusion of SES. More interesting, however, is that the variance of the intercept parameter in the random-effects regression model is only half as large as the variance of the intercept in the random-effects ANOVA model. The reason for this is that even though SES is a student-level predictor, it also carries information about differences between schools – the average SES within some schools is higher than others. This variation in the average SES of students from different schools is accounting for a great deal of variation in the school means, hence the much lower value of $\hat{\tau}_{00}$ in the random-effects regression.

We will discuss ways to separate within-group and between-group effects of Level 1 predictors in the subsequent section.

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
Intercept	12.6575	0.1880	159	67.34	<.0001	0.05
student_ses	2.3903	0.1057	7024	22.61	<.0001	0.05

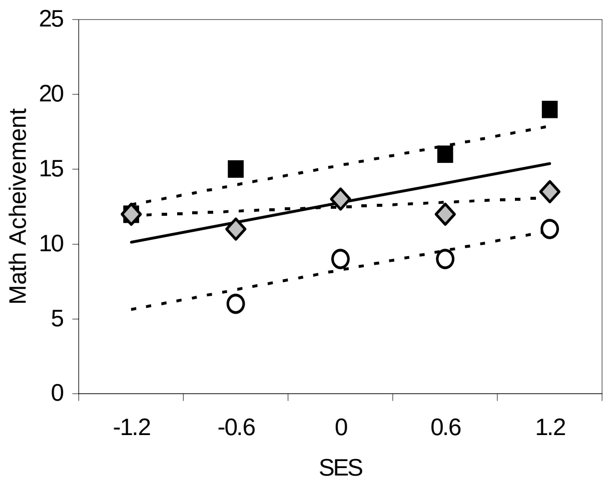
Solution for Fixed Effects

Effect	Lower	Upper
Intercept	12.2863	13.0287
student_ses	2.1830	2.5975

The average intercept across schools is estimated to be $\hat{\gamma}_{00} = 12.66$. However, the significant variance component for the intercept indicates that this value varies considerably across schools. The effect of student SES is now estimated as $\hat{\gamma}_{10} = 2.39$. This value is somewhat smaller than the one estimated in the fixed-effects regression model, again related to the confounding of within- and between-school differences on SES in the multilevel analysis.

Random Slopes

Importantly, the SES effect estimated in the preceding model is assumed to be constant across all schools as there is no corresponding random effect associated with SES. However, we might have reason to believe that the effect of SES varies across schools. Recent educational policy has emphasized the need to promote *equity* within schools, bringing the achievement of underprivileged students to the same high level often observed for more affluent students. Some schools may emphasize this more than others, leading to differences in the effect of SES. In general, more equitable schools would be those showing weaker SES effects. For instance, the observations added to the plot belong to a school showing high equity. To accommodate such differences between schools in the effect of SES, we need to incorporate random slopes into our mixed model.



Random Slopes Model

Level 1 Equation:

$$\text{Math}_{ij} = \beta_{0j} + \beta_{1j}\text{SES}_{ij} + r_{ij} \qquad r_{ij} \sim N(0,\sigma^2)$$

Level 2 Equations:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \qquad \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$$

Reduced-Form Equation:

$$\begin{aligned} \text{Math}_{ij} &= \left(\gamma_{00} + u_{0j}\right) + \left(\gamma_{10} + u_{1j}\right)\text{SES}_{ij} + r_{ij} \\ &= \underbrace{\left(\gamma_{00} + \gamma_{10}\text{SES}_{ij}\right)}_{\text{Fixed-Effects}} + \underbrace{\left(u_{0j} + u_{1j}\text{SES}_{ij}\right)}_{\text{Random-Effects}} + r_{ij} \end{aligned}$$

The PROC MIXED code for this model is then specified as follows:

```
proc mixed data=mixed.hsb method=reml cl covtest;
   model student_mathach = student_ses / solution alpha=.05 ddfm=bw
                            outpred=predicted;
   random intercept student_ses / subject=school_id type=un g gcorr;
run;
```

Notice that **student_ses** has been added to the RANDOM statement. Further, the option TYPE=UN allows the random intercepts and slopes to covary (so that τ_{10} is estimated and not inadvertently fixed to zero). We will discuss the OUTPRED option shortly.

PROC MIXED Output:

```
                            Model Information

          Data Set                    MIXED.HSB
          Dependent Variable          student_mathach
          Covariance Structure        Unstructured
          Subject Effect              school_ID
          Estimation Method           REML
          Residual Variance Method    Profile
          Fixed Effects SE Method     Model-Based
          Degrees of Freedom Method   Between-Within

                                Dimensions

                    Covariance Parameters             4
                    Columns in X                      2
                    Columns in Z Per Subject          2
                    Subjects                        160
                    Max Obs Per Subject              67
```

The Covariance Structure (our **T**) matrix is now listed as unstructured, we have two columns in Z (two random effects), and so we have a total of four covariance parameters, σ^2, τ_{00}, τ_{10}, and τ_{11} to estimate.

```
                            Estimated G Matrix

           Row    Effect         Subject        Col1        Col2

             1    Intercept            1      4.8278     -0.1547
             2    student_ses          1     -0.1547      0.4127


                      Estimated G Correlation Matrix

           Row    Effect         Subject        Col1        Col2

             1    Intercept            1      1.0000     -0.1096
             2    student_ses          1     -0.1096      1.0000

                      Covariance Parameter Estimates

                                           Standard        Z
 Cov Parm    Subject     Estimate       Error     Value        Pr Z    Alpha

 UN(1,1)     school_ID     4.8278      0.6719      7.18      <.0001     0.05
```

UN(2,1)	school_ID	-0.1547	0.2988	-0.52	0.6046	0.05
UN(2,2)	school_ID	0.4127	0.2350	1.76	0.0395	0.05
Residual		36.8304	0.6293	58.52	<.0001	0.05

Cov Parm	Subject	Lower	Upper
UN(1,1)	school_ID	3.7406	6.4716
UN(2,1)	school_ID	-0.7403	0.4308
UN(2,2)	school_ID	0.1730	1.9418
Residual		35.6274	38.0956

Our Level 2 covariance parameter estimates are listed by the row and column position in the unstructured covariance matrix for the random effects. Our estimated **T** matrix (referred to as **G** in SAS) is

$$\mathbf{T} = \begin{pmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{pmatrix} = \begin{pmatrix} \text{un}(1,1) & \\ \text{un}(2,1) & \text{un}(2,2) \end{pmatrix} = \begin{pmatrix} 4.8278 & \\ -.1547 & .4127 \end{pmatrix}$$

The row 1, column 1 position, UN(1,1), is $\hat{\tau}_{00}$, the row 2, column 1 position, UN(2,1), is $\hat{\tau}_{10}$, and the row 2, column 2 position, UN(2,2), is $\hat{\tau}_{11}$.

Examining the actual estimates, we see that the slope variance is statistically significant, indicating that there are in fact differences in equity across schools. In some schools SES is a weaker predictor of math achievement scores than in others.

Is it the case that schools with weak SES effects are also high-performing, consistent with the notion that the scores of impoverished students have been brought up to the level of affluent students? The correlation of -.11 between the random intercepts and slopes indicates that schools with higher intercepts do indeed have lower (less positive) slopes for SES; however, the confidence interval for the corresponding covariance estimate runs from -.74 to .43, indicating that this relationship is not significant.

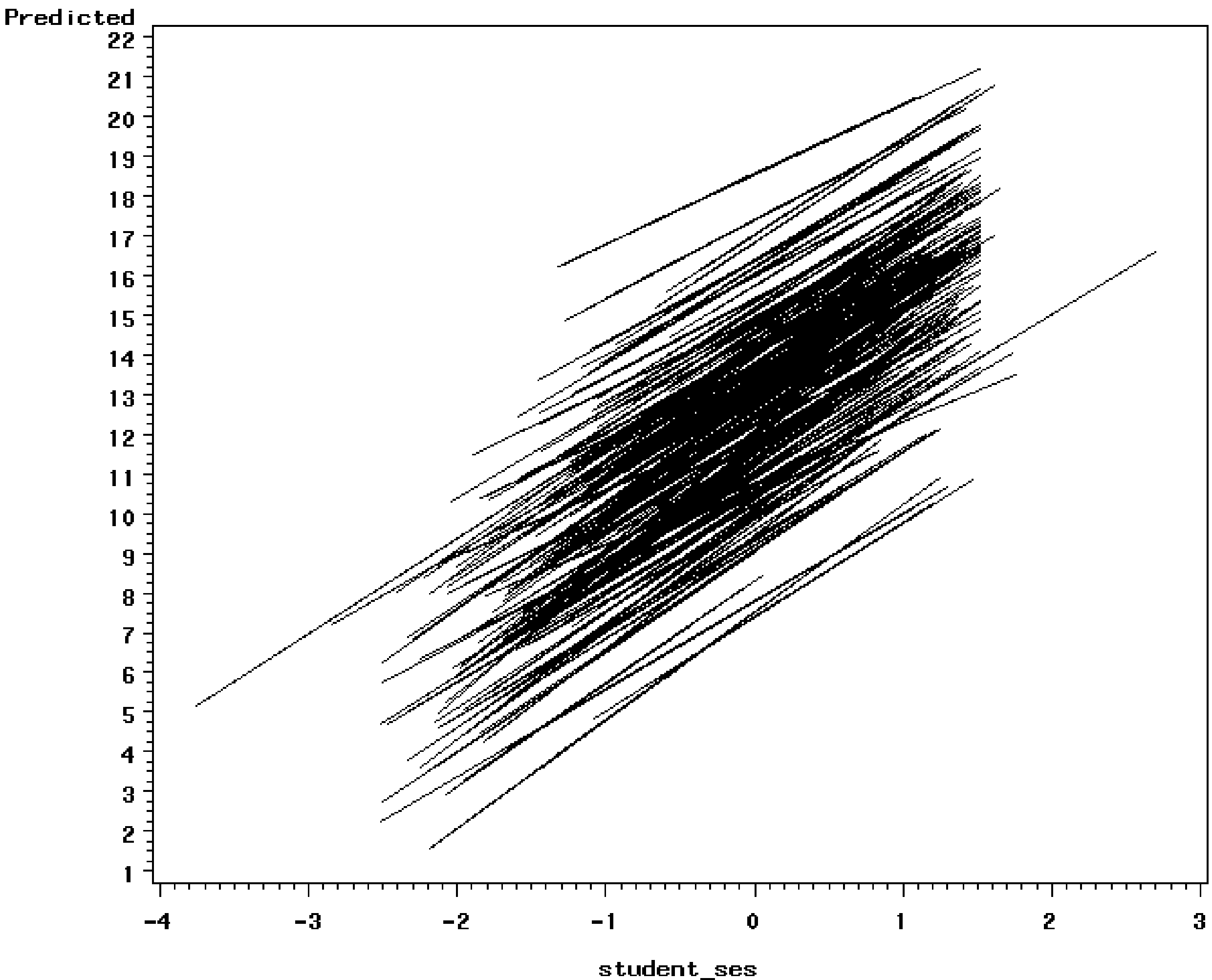
Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha
Intercept	12.6651	0.1898	159	66.72	<.0001	0.05
student_ses	2.3938	0.1181	7024	20.27	<.0001	0.05

The solution for the fixed effects is interpreted as before, with the caveat that the estimate given here for **`student_ses`** is an average effect. We know from the significance of the variance component for this variable that the SES effect varies from this average across schools.

To visualize these differences we will now make use of the OUTPRED option. This option created an output data set of predicted values for each student's math achievement within each school. These predicted values take into account school differences in intercepts and SES slopes through the use of empirical Bayes' estimates. Empirical Bayes' estimates are estimates for each school intercept and slope that are obtained subsequent to fitting the model (they are not generated to fit the model itself). These estimates are shrunken, meaning that they are pulled toward the fixed effects and thus do not represent the full extent of variability across schools. Nevertheless, plotting the predicted values obtained by the OUTPRED option provides a useful visualization of the patterns in the data.

```
goptions reset=all hsize=7 vsize=6;
symbol i=join color=black repeat=160;
proc gplot data=predicted;
   plot pred*student_ses=school_id/nolegend;
run; quit;
```



This plot reflects that, although all school-level regression lines are positive, there appears to be much variability among both the intercepts and the slopes across the schools. In later models, we will attempt to incorporate additional predictors in an attempt to account for some of this variability.

1.6 Chapter Summary

Summary of Introduction to Multilevel Models

- Dependence is often a feature of data structures, particularly if data is hierarchical or longitudinal.
- Dependence presents a problem to traditional statistical methods, which assume that errors of prediction are uncorrelated.
- Several methods are available for modeling dependent data structures; multilevel models have a number of important advantages.
- The most basic multilevel model is the random-effects ANOVA model, and this provides a measure of the degree of nesting via the ICC.
- Random-effects regression allows the effects of Level 1 predictors to vary across groups.

40

In this chapter, we introduced the notion of nesting and the dependence that it produces in data structures. We noted that dependence typically arises due either to a hierarchical data structure (for example, individuals nested within groups) or a longitudinal data structure (for example, repeated measures collected for each person).

We noted that dependence is a critical problem for conventional methods of analysis. The crux of the problem is that methods that assume independence assume that each observation constitutes a unique piece of information, but less unique information is actually available when observations are correlated. Assuming independence for such data causes conventional models to overstate precision and power, producing overly optimistic standard errors and test statistics.

Options to more appropriately model dependent data structures include these:

- A conventional model with corrected standard errors and test statistics
- The fixed-effects approach in which sources of dependence are covaried out of the model
- The multilevel model (or random-effects approach), which explicitly treats both individuals and groups as two levels of sampling

Advantages of multilevel models include parsimony, broader inference, conformity to our conceptual models, and the ability to separate effects at each level of the model (e.g., individual and contextual).

Finally, this chapter has introduced some basic multilevel models. The random-effects ANOVA is a useful initial model that provides an estimate of the intra-class correlation (ICC) and unconditional estimates of the variance components in the model. The random-effects regression model adds one or more Level 1 predictors to the multilevel model. These predictors can have either fixed or random slopes. If the slope is fixed, this presumes that the predictor has the same effect in every group. If it is random, the effect of the predictor varies across groups.

1.7 Exercises

[The data for this exercise was artificially generated.]

For this set of exercises we will use a simulated data set designed to reflect substantive questions and empirical characteristics that might be encountered in a business or marketing research setting. The data set contains measures on N=2859 employees nested within J=100 managers within a multi-national corporation. There are between 15 and 50 employees nested within any given manager. The dependent variable is a continuous measure of current annual salary measured in units of thousands of dollars. The variables of interest are listed here:

emp_id	unique identifier of each employee ranging from 1 to 2859
man_id	unique identifier of each manager ranging from 1 to 100
man_prf_grn	grand-mean centered continuous measure of the job performance of the manager; higher scores reflect higher levels of performance
man_mba	a dichotomous indicator reflecting whether the manager does or does not hold an MBA degree (0=no MBA; 1=MBA)
emp_prf	uncentered continuous measure of the job performance of the employee; higher scores reflect higher levels of performance
emp_prf_grp	group-mean centered measure of employee performance
emp_prf_grn	grand-mean centered measure of employee performance
emp_prf_mean	mean of employee performance within each group
emp_rep	uncentered continuous measure of the difficulty of replacing a given employee; higher scores reflect greater difficulty in replacement (thus reflecting unique job skills)
emp_rep_grp	group-mean centered measure of employee replaceability
emp_rep_grn	grand-mean centered measure of employee replaceability
emp_rep_mean	mean of employee replaceability within each group
emp_college	a dichotomous indicator reflecting whether the employee does or does not hold a college degree (for example, BA/BS; 0=no degree; 1=degree)
emp_salary	uncentered continuous measure of employee annual salary measured in thousands of dollars (for example, 34.754 is $34,754).

The data is stored in a SAS data set named **`mixed.sim`**.

1. Use PROC PRINT to examine employee ID (**`emp_id`**), manager ID (**`man_id`**), employee salary (**`emp_salary`**), and grand-mean centered employee performance (**`emp_prf_grn`**) for the first 50 employees.

2. Use PROC UNIVARIATE to examine the univariate statistics for employee salary. What is the mean, median, standard deviation, and range? Plot a histogram of this measure. Are there any potential outlying observations? Does salary appear to be normally distributed?

3. Use PROC REG to regress employee salary on the grand-mean centered measure of employee performance (**`emp_prf_grn`**). Summarize the results. Given the nesting of employee within manager, what assumptions might be violated here? If these are violated, what might be the impact on the regression results?

4. Expand the regression model to also include the grand-mean centered measure of employee replaceability (**`emp_rep_grn`**). Summarize the results. Why does the effect of employee performance not change much in the second model compared to the first? (Use PROC CORR to evaluate the relationship between the two predictors.)

5. Use PROC MEANS to compute the mean employee salary within each of the 100 managers. Is there variability in the number of employees nested within each manager? Does there appear to be meaningful variability among the group means?

6. Write out the Level 1, Level 2, and reduced-form equation for the random-effects ANOVA model with employee salary (**`emp_salary`**) in which employees are nested within manager (**`man_id`**). Estimate this model using PROC MIXED to examine whether there is evidence of dependence among employees nested within managers. Compute the intraclass correlation and provide a brief interpretation of this value. Provide brief interpretations for all of the key values in the MIXED output.

7. Add to the model from #6 the **grand-mean centered** measure of employee performance (**`emp_rep_grn`**) as the sole Level 1 predictor. Do not estimate any random effects (that is, remove the random intercept). Briefly interpret these results. How would you expect these results to compare to the same model estimated in PROC REG in #3?

8. To the model from #7, add a random intercept (but not slope). Write out the Level 1, Level 2, and reduced-form expression and estimate this in PROC MIXED. Briefly summarize your findings.

9. To the model from #8 add a random slope component as well. Is there evidence for whether the random slope component is needed in the model? Use the OUTPRED option to generate a plot of the variability in the relation of **`emp_prf_grn`** to **`emp_salary`**. Briefly summarize your findings.