Introduction to Hierarchical Linear Models

Course Notes

Introduction to Hierarchical Linear Models Course Notes was developed by Mike Patetta.

Introduction to Hierarchical Linear Models Course Notes

Book code E71539, course code ATEHLM, prepared date 19Jul2019. ATEHLM_001

Table of Contents

To learn more…

For information about other courses in the curriculum, contact the SAS Education Division at 1-800-333-7660, or send e-mail to training@sas.com. You can also find this information on the web at http://support.sas.com/training/ as well as in the Training Course Catalog.

For a list of SAS books (including e-books) that relate to the topics covered in this course notes, visit https://www.sas.com/sas/books.html or call 1-800-727-0025. US customers receive free shipping to US addresses.

Lesson 1 Introduction to Hierarchical Linear Models

1.1 Introduction to Hierarchical Linear Models

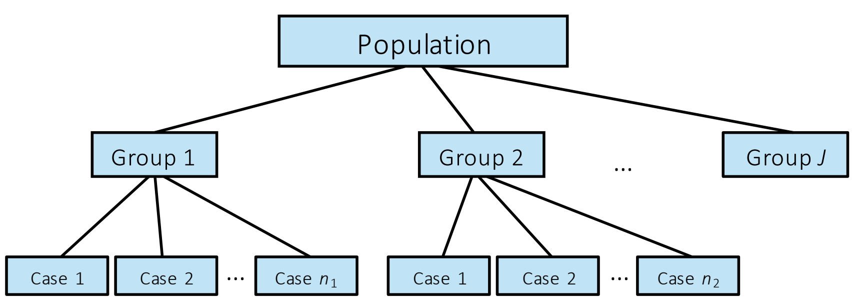
Nested Data

- Multilevel models are used when you have nested data.
- Nested data typically comes in one of two forms: hierarchical or longitudinal.

2

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).



3

Unintentional Sources of Nesting

- Nesting might also occur even when it is not an explicit part of the study design and thus is unintentional.
- Consider the following examples:
  - respondents nested within an interviewer
  - homeless adolescents nested within social service sectors
  - multiple specimens nested within a laboratory

4

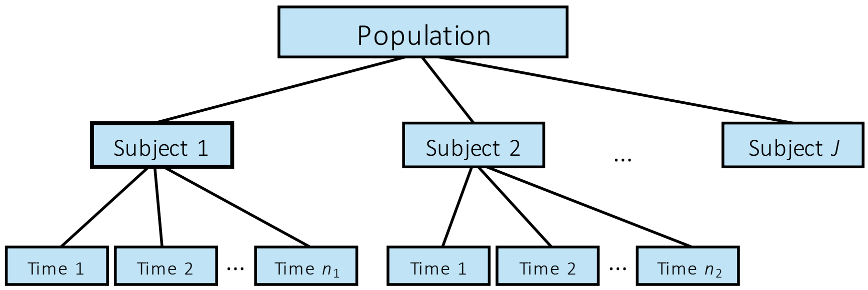
Dependence in Hierarchical Data

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

5

Longitudinal Data Structures

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



6

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.

7

The High School and Beyond Data Set

- To demonstrate hierarchical linear models, we will use the High School and Beyond (**hsb**) survey data set.
- The sample consists of 7185 students from 160 schools.
- The continuous response variable is student's mathematical ability.
- Level 1 predictors are student's socioeconomic status, gender, and race.
- Level 2 predictors are school size, whether it is a public school, and the disciplinary climate of the school.

8

Assumptions of the OLS Regression Model

The ordinary regression model makes a number of assumptions, including the following:

- The mean of errors is zero.
- The predictors are uncorrelated with errors.
- Predictor levels are fixed.
- Errors are normally distributed (for inference).
- Errors are homoscedastic (have constant variance).
- Errors are independent (uncorrelated).

Note: The last three are reflected in the term *normal-iid*. This means that the errors are normal, independent, and identically distributed.

9

Potential Violations of Assumptions

- Of these assumptions, the assumption of independence is particularly dubious given the nesting of students within schools.
- Homoscedasticity might also be violated, if the effect of socioeconomic status (SES) actually varies among schools.

10

Three Key Problems When Violating Independence

If the observations are positively correlated:

1. *F* statistics tend to be too large.
   - You are likely to overestimate the significance of the model as a whole.
2. Standard error estimates tend to be too small.
   - You are likely to overestimate the significance of specific regression coefficients.
3. Ignoring the hierarchical structure of data severely limits the ability to model within and between group effects that might be of key interest.

11

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. You think that 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 potentially 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.

Classical Approach to Multilevel Data

A classical approach to modeling a multilevel problem, such as students nested within schools, is to include schools as a fixed effect in the model.

This has two major disadvantages:

1. explosion in the number of parameters to estimate
2. narrow scope of inference

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.
- You can write the regression model as follows:

$$\text{Math}_{ij} = b_{0j} + b_1\text{SES}_{ij} + \varepsilon_{ij}$$

where you assume a particular distribution for b_{0j}, as you customarily do for ε_{ij}.

14

Advantages of Multilevel Models

Multilevel models have the following advantages:

- are 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 you to examine the effects of individual-level and group-level influences simultaneously
- enable you to estimate contextual effects

15

1.2 The Random-Effects ANOVA Model

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 among the groups.
- Like the fixed-effects ANOVA, you can decompose the variance of the observed variable into within-group and between-group components.

16

MIXED Procedure

General form of the MIXED procedure:

```
PROC MIXED DATA=SAS-data-set <options>;
      CLASS variables;
      MODEL response = <fixed-effects> </options>;
      RANDOM random-effects </options>;
      REPEATED <repeated-effect> </options>;
RUN;
```

17

The Random-Effects ANOVA Model

(micro) Level 1:

$$y_{ij} = b_{0j} + \varepsilon_{ij}$$

(macro) Level 2:

$$b_{0j} = \beta_{00} + b*_{0j}$$

Reduced Form:

$$y_{ij} = \beta_{00} + b*_{0j} + \varepsilon_{ij}$$

18

The Variance Components of the Model

- The reduced form equation is shown below:

$$y_{ij} = \beta_{00} + b*_{0j} + \varepsilon_{ij}$$

- It includes one fixed effect (the grand mean) and two random components (the residual error at Level 1 and the error at Level 2).
- Assume that these errors are normally distributed and uncorrelated with one another:

$$\varepsilon_{ij} \sim N(0,\sigma^2) \quad b*_{0j} \sim N(0,\sigma^2{}_{00})$$

- This implies the following:

$$\mathrm{E}(y_{ij}) = (\beta_{00})$$

$$V(y_{ij}) = V\left(b*_{0j} + \varepsilon_{ij}\right) = V(b*_{0j}) + V(\varepsilon_{ij}) = \sigma^2{}_{00} + \sigma^2$$

19

The Intraclass Correlation

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

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

- 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: The general linear model assumes ICC=0.

20

Estimating Multilevel Models in SAS

- Linear multilevel models, such as the random-effects ANOVA model, can be estimated in SAS with the MIXED procedure.
- The reduced form equation is used to define the model in PROC 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 PROC MIXED is to first write the Level 1 and Level 2 equations and then construct the reduced form equation from these expressions.

21

Fitting a Random-effects ANOVA Model with PROC MIXED

This demonstration calculates the intraclass correlation between students within a school to determine the degree of dependence that is present in the data. Examine the partial contents of the data set using PROC PRINT.

```
proc print data=mixed.hsb (obs=15);
   var school_id student_id student_ses school_disclim
       student_mathach;
run;
```

Line Listing of High School and Beyond Data Set

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
6	1224	6	0.022	1.597	4.583
7	1224	7	-0.618	1.597	-2.832
8	1224	8	-0.998	1.597	0.523
9	1224	9	-0.888	1.597	1.527
10	1224	10	-0.458	1.597	21.521
11	1224	11	-1.448	1.597	9.475
12	1224	12	-0.658	1.597	16.057
13	1224	13	-0.468	1.597	21.178
14	1224	14	-0.988	1.597	20.178
15	1224	15	0.332	1.597	20.349

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 students within a school, but vary across schools.

Next sort the data by **school_id** so that the data are ordered properly for the MIXED analyses. You do not need to manually sort the data if you use a CLASS statement in PROC MIXED. However, each time that PROC MIXED encounters the CLASS statement, the data are re-sorted. If the data are already sorted, omitting the CLASS statement can increase computational efficiency.

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

Specify a random ANOVA model:

```
proc mixed data=mixed.hsb cl covtest;
   model student_mathach = / solution ddfm=bw;
   random intercept/subject=school_id v vcorr;
   title 'Math Achievement: Random-effect ANOVA';
run;
```

The options selected for the PROC MIXED statement are **cl** and **covtest**. The first of these requests confidence limits for the variance component estimates, and the second requests asymptotic standard errors and Wald tests for covariance parameters. It is important to note that both of these are based on large-sample approximations and require a fairly large sample (number of schools in this case) to be considered useful. The **hsb** data set contains data for 160 schools. The **covtest** generally requires larger samples than the **cl** option and might not be appropriate here.

The MODEL statement specifies the fixed effects for the model. In this case, there are no predictors with fixed effects, only 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.

DDFM=BW — requests that the degrees of freedom for testing the fixed effects be computed using the BETWEEN/WITHIN method. This method is what is typically used for multilevel models, and it is appropriate in a large sample. Better methods are available for small samples, including Satterthwaite (DDFM=SATTER) and improved Kenward-Roger (DDFM=KR2). KR2 is an updated version of KR that performs better for nonlinear covariance structures.

The RANDOM statement specifies the random effects in the model. In this case, the INTERCEPT keyword specifies that the intercept is to be a random effect. Selected options for the RANDOM statement include the following:

SUBJECT= — indicates the nesting structure of the data. Defining **SUBJECT=school_id** tells PROC MIXED that the intercept is to vary randomly across schools. This variable is typically also specified in the CLASS statement. However, if only one SUBJECT= variable is used, and if the data are sorted by the SUBJECT= variable, then the SUBJECT= variable can be omitted from the CLASS statement to improve computational speed and memory usage.

V — requests that the covariance matrix among observations (Level 1 observations) be displayed.

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

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

The Model Information section is useful for making sure that you specified the model as intended. *Variance Components* is the default structure for the covariance matrix of the random effects at Level 2. It specifies that the random effects are independent. In this case, this assumption is acceptable because you 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 table provides information about the size of the model and data set. The two covariance parameters correspond to σ^2_{00} and σ^2. In PROC MIXED, the X matrix is the design matrix for the fixed effects. Because the model includes only one fixed effect, the intercept β_{00}, this is listed as one. The Z matrix is the design matrix for the random effects at Level 2. There is a random intercept, so this is also listed as one. Finally, PROC MIXED reports 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 converged. ***If the model does not converge, do not interpret the model estimates.*** You might need to increase the number of iterations that PROC MIXED performs, use different start values (written by the PARMS statement), or there might be a problem with your model.

In this model, the V and VCORR options provide the estimate of the intraclass covariance and correlation matrices for the first group, respectively. The V and VCORR matrices are the same across all groups, so you need to consider only one. Here are the estimates for the first subject (partial output):

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

Partial Output

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 = 0.1803. Notice 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 (for example, compound symmetry).

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	Lower	Upper
Intercept	school_ID	8.6097	1.0778	7.99	<.0001	0.05	6.8339	11.1843
Residual		39.1487	0.6607	59.26	<.0001	0.05	37.8855	40.4765

The Covariance Parameter Estimates section provides estimates of the variance components, which are $\hat{\sigma}_{00}^2 = 8.61$ and $\hat{\sigma}^2 = 39.15$.

Thus, the estimate of the ICC is $\dfrac{\hat{\sigma}_{00}^2}{\hat{\sigma}_{00}^2 + \hat{\sigma}^2} = \dfrac{8.61}{8.61 + 39.15} = 0.18$.

In other words, 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 0.18. This is, of course, the same value that was presented in the VCORR matrix.

In addition, you 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, you can see that the precision is greater for the variance component at Level 1 than for the variance component at Level 2. This result is quite typical for multilevel models. These confidence intervals never include zero, so in practice, very small values for the lower bound might indicate that the variance component is unnecessary. Alternatively, you can use the tests supplied by COVTEST to assess the variance components. In this case, both are statistically significant.

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

Note: The COVTEST option produces Wald z-tests and p-values for all variance and covariance parameter estimates. However, these tests assume an 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, which makes these tests inaccurate.

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

There is no direct test of the ICC, but notice that the ICC is zero when σ^2_{00} is zero, so typically the test of $\hat{\sigma}^2_{00}$ is used as a proxy for a test of the estimated ICC.

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.6370	0.2443	159	51.72	<.0001

The fixed intercept estimate, which is the estimate of the average group mean, is $\hat{\beta}_{00} = 12.64$. This value differs slightly from the grand mean. If the groups differ in size, as in this case, the average of the estimated group means does not necessarily equal the grand mean estimate.

End of Demonstration

1.3 Random-Effects Regression Model

One Predictor: Fixed Intercept and Fixed Slope

Level 1 Model:

$$y_{ij} = b_{0j} + b_{1j}x_{ij} + \varepsilon_{ij}$$

Level 2 Model:

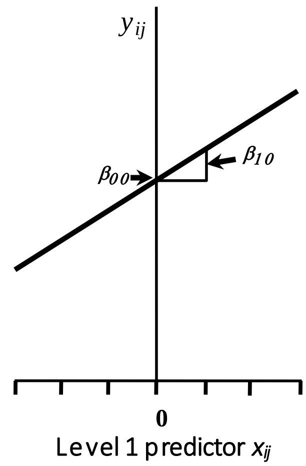
$$b_{0j} = \beta_{00}$$

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

Reduced-Form Model:

$$y_{ij} = \beta_{00} + \beta_{10}x_{ij} + \varepsilon_{ij}$$

Fixed Intercept, Fixed Slope



23

One Predictor: Fixed Intercept and Fixed Slope

Level 1 Model:

$$y_{ij} = b_{0j} + b_{1j}x_{ij} + \varepsilon_{ij}$$

Level 2 Model:

$$b_{0j} = \beta_{00}$$

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

Reduced-Form Model:

$$y_{ij} = \beta_{00} + \beta_{10}x_{ij} + \varepsilon_{ij}$$

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

```
proc mixed data=mydata cl covtest;
   model y = x / solution;
run;
```

24

One Predictor: Random Intercept and Fixed Slope

Level 1 Model:

$$y_{ij} = b_{0j} + b_{1j}x_{ij} + \varepsilon_{ij}$$

Level 2 Model:

$$b_{0j} = \beta_{00} + b^*_{0j}$$

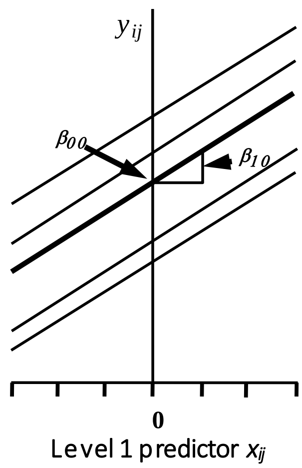
$$b_{1j} = \beta_{10}$$

Reduced-Form Model:

$$y_{ij} = \left(\beta_{00} + b^*_{0j}\right) + \beta_{10}x_{ij} + \varepsilon_{ij}$$

$$= \left(\beta_{00} + \beta_{10}x_{ij}\right) + b^*_{0j} + \varepsilon_{ij}$$

Random Intercept, Fixed Slope



One Predictor: Random Intercept and Fixed Slope

Level 1 Model:

$$y_{ij} = b_{0j} + b_{1j}x_{ij} + \varepsilon_{ij}$$

Level 2 Model:

$$b_{0j} = \beta_{00} + b^*_{0j}$$

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

Reduced-Form Model:

$$y_{ij} = \left(\beta_{00} + b^*_{0j}\right) + \beta_{10}x_{ij} + \varepsilon_{ij}$$

$$= \left(\beta_{00} + \beta_{10}x_{ij}\right) + b^*_{0j} + \varepsilon_{ij}$$

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

$$b^*_{0j} \sim N(0, \sigma^2_{00})$$

```
proc mixed data=mydata cl covtest;
   model y=x / solution;
   random intercept / subject=L2_ID;
run;
```

One Predictor: Random Intercept and Random Slope

Level 1 Model:

$$y_{ij} = b_{0j} + b_{1j}x_{ij} + \varepsilon_{ij}$$

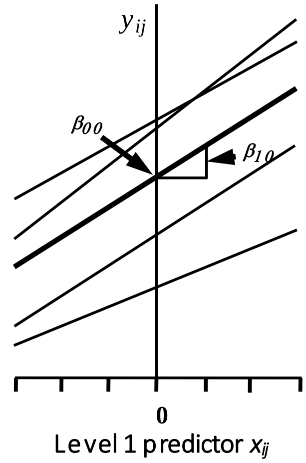
Level 2 Model:

$$b_{0j} = \beta_{00} + b^*_{0j}$$
$$b_{1j} = \beta_{10} + b^*_{1j}$$

Reduced-Form Model:

$$y_{ij} = \left(\beta_{00} + b^*_{0j}\right) + \left(\beta_{10} + b^*_{1j}\right)x_{ij} + \varepsilon_{ij}$$
$$= \left(\beta_{00} + \beta_{10}x_{ij}\right) + \left(b^*_{0j} + b^*_{1j}\, x_{ij}\right) + \varepsilon_{ij}$$

Random Intercept, Random Slope



27

Covariance Matrices of Residuals and Random Effects

$$\text{var}(\varepsilon_{ij}) = \sigma^2$$

Level 1

$$\text{var}(b^*_{0j}) = \sigma^2_{00}$$
$$\text{var}(b^*_{1j}) = \sigma^2_{11}$$
$$\text{cov}(b^*_{0j}, b^*_{1j}) = \sigma_{10} = \sigma_{01}$$

Level 2

$$\mathbf{G} = \begin{pmatrix} \sigma^2_{00} & \sigma_{01} \\ \sigma_{10} & \sigma^2_{11} \end{pmatrix}$$

Covariance matrix

28

One Predictor: Random Intercept and Random Slope

Level 1 Model:

$$y_{ij} = b_{0j} + b_{1j}x_{ij} + \varepsilon_{ij}$$

Level 2 Model:

$$b_{0j} = \beta_{00} + b^*_{0j}$$
$$b_{1j} = \beta_{10} + b^*_{1j}$$

Reduced-Form Model:

$$y_{ij} = \left(\beta_{00} + b^*_{0j}\right) + \left(\beta_{10} + b^*_{1j}\right)x_{ij} + \varepsilon_{ij}$$
$$= \left(\beta_{00} + \beta_{10}x_{ij}\right) + \left(b^*_{0j} + b^*_{1j}\,x_{ij}\right) + \varepsilon_{ij}$$

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

$$\begin{bmatrix} b^*_{0j} \\ b^*_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{00} & \\ \sigma_{10} & \sigma^2_{11} \end{bmatrix}\right)$$

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

29

Fitting a Random-Effects Regression with PROC MIXED

For this demonstration, use the **hsb** data set to fit the three regression models described in the slides and compare their substantive implications. For these analyses, the dependent variable is math achievement, **student_mathach**, and the predictor is the socioeconomic status of the student's family, **student_ses**. The nesting of students within schools is indicated by **school_ID**.

Simple Regression Model

Begin with the simple regression model with a fixed intercept and slope:

Level 1 Equation:

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

Level 2 Equations:

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

Reduced-Form Equation:

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

To enter this model into the MIXED procedure, write the following code:

```
proc mixed data=mixed.hsb cl covtest;
   model student_mathach = student_ses / solution;
   title 'Math Achievement: Regression Model, SES';
run;
```

The results from this model provide a useful baseline for judging subsequent models.

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 you are assuming independence of observations. This indicates that the model reduces to an ordinary least squares model.

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, a column for the intercept and a column for **student_ses**. Z is the design matrix for the random effects, which in this case is empty, because there are no random effects. 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

The estimate of the Level 1 residual error variance is $\hat{\sigma}^2 = 41.16$.

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.7474	0.07569	7183	168.42	<.0001
student_ses	3.1839	0.09712	7183	32.78	<.0001

Because **student_ses** is mean-centered (has a mean of zero), you can interpret these estimates as follows:

The estimate $\hat{\beta}_{00} = 12.75$ is the expected math achievement of a student from an average SES family; and the estimate $\hat{\beta}_{10} = 3.18$ indicates the expected increase in math achievement per one-unit increase on the SES index.

One important feature of this model 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 come from.

The Random Intercepts Model

The mixed model assumes that these school-specific intercepts are normally distributed around the ***average*** regression line. Further, it assumes that the residual errors for the students are independently and normally distributed around their ***group*** regression line. These seem to be much more reasonable assumptions. The model equations are shown here.

Level 1 Equation:

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

Level 2 Equations:

$$b_{0j} = \beta_{00} + b^*_{0j} \qquad b^*_{0j} \sim N(0, \sigma^2_{00})$$
$$b_{1j} = \beta_{10}$$

Reduced-Form Equation:

$$\text{Math}_{ij} = \left(\beta_{00} + b^*_{0j}\right) + \beta_{10}\text{SES}_{ij} + \varepsilon_{ij}$$
$$= \underbrace{\left(\beta_{00} + \beta_{10}\text{SES}_{ij}\right)}_{\text{Fixed-Effects}} + \underbrace{b^*_{0j}}_{\text{Random-Effect}} + \varepsilon_{ij}$$

In PROC MIXED, specify the following:

```
proc mixed data=mixed.hsb cl covtest;
   model student_mathach = student_ses / solution ddfm=bw;
   random intercept / subject=school_id;
  title 'Math Achievement: Random-effect ANOVA';
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: The covariance structure is variance components, and you now have two covariance parameters, σ^2 and σ^2_{00}. There are still two fixed-effects predictors, so the columns in **X** are unchanged, but the number of columns in **Z** is now one, which reflects the random intercept.

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

Covariance Parameter Estimates								
Cov Parm	**Subject**	**Estimate**	**Standard Error**	**Z Value**	**Pr > Z**	**Alpha**	**Lower**	**Upper**
Intercept	school_ID	4.7665	0.6549	7.28	<.0001	0.05	3.7045	6.3636
Residual		37.0346	0.6254	59.22	<.0001	0.05	35.8388	38.2916

The estimates of the variance components are $\hat{\sigma}_{00}^2 = 4.77$ 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 resulted in a reduction in the Level 1 residual variance (from 41 to 37). This is to be expected, because 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 from the last demonstration. In the random-effects ANOVA, $\hat{\sigma}_{00}^2 = 8.61$ 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, you 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.

However, it is more interesting that the variance of the intercept parameter in the random-effects regression model (4.77) is only about half as large as the variance of the intercept in the random-effects ANOVA model (8.61). The reason for this is that even though SES is a student-level predictor, it also carries information about differences among schools. That is, the average SES within some schools is higher than others. This variation in the average SES of students from different schools accounts for a great deal of variation in the school means; hence, the much lower value of $\hat{\sigma}_{00}^2$ in the random-effects regression.

Solution for Fixed Effects					
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	12.6575	0.1880	159	67.34	<.0001
student_ses	2.3903	0.1057	7024	22.61	<.0001

The average intercept across schools is estimated to be $\hat{\beta}_{00} = 12.66$. 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 $\hat{\beta}_{10} = 2.39$. This value is somewhat smaller than the one estimated in the fixed-effects regression model, again related to the confounding of within-school and between-school differences on SES in the multilevel analysis.

Random Slopes

The SES effect estimated in the preceding model is assumed to be constant across all schools because there is no corresponding random effect associated with SES. However, you might have reason to believe that the effect of SES varies across schools. Recent educational policy emphasized the need to promote ***equity*** within schools to bring the achievement level of underprivileged students to the same high level often observed for more affluent students. Some schools might emphasize this more than others, which can lead to differences in the effect of SES. In general, more equitable schools are those showing weaker

SES effects. For example, the observations added to the plot belong to a school showing high equity. To accommodate such differences between schools in the effect of SES, you need to incorporate random slopes into the mixed model.

Random Slopes Model

Level 1 Equation:

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

Level 2 Equations:

$$b_{0j} = \beta_{00} + b*_{0j} \qquad \begin{pmatrix} b*_{0j} \\ b*_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{00} & \\ \sigma_{10} & \sigma^2_{11} \end{pmatrix}\right]$$

$$b_{1j} = \beta_{10} + b*_{1j}$$

Reduced-Form Equation:

$$\begin{aligned} \text{Math}_{ij} &= \left(\beta_{00} + b*_{0j}\right) + \left(\beta_{10} + b*_{1j}\right)\text{SES}_{ij} + \varepsilon_{ij} \\ &= \underbrace{\left(\beta_{00} + \beta_{10}\text{SES}_{ij}\right)}_{\text{Fixed-Effects}} + \underbrace{\left(b*_{0j} + b*_{1j}\,\text{SES}_{ij}\right)}_{\text{Random-Effects}} + \varepsilon_{ij} \end{aligned}$$

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

```
proc mixed data=mixed.hsb cl covtest;
   model student_mathach = student_ses / solution ddfm=bw;
   random intercept student_ses / subject=school_id type=un g
          gcorr;
   title 'Math Achievement: Random Intercept and Slope';
run;
```

Notice that **student_ses** was added to the RANDOM statement. Further, the TYPE=UN option enables the random intercepts and slopes to covary (so that σ_{10} is estimated and not fixed to zero).

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 matrix of random effects (**G**) is unstructured. There are two columns in Z (two random effects: the intercept and the slope). In total, there are four covariance parameters to estimate: σ^2, σ^2_{00}, σ_{10}, and σ^2_{11}.

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								
Cov Parm	**Subject**	**Estimate**	**Standard Error**	**Z Value**	**Pr Z**	**Alpha**	**Lower**	**Upper**
UN(1,1)	school_ID	4.8278	0.6719	7.18	<.0001	0.05	3.7406	6.4716
UN(2,1)	school_ID	-0.1547	0.2988	-0.52	0.6046	0.05	-0.7403	0.4308
UN(2,2)	school_ID	0.4127	0.2350	1.76	0.0395	0.05	0.1730	1.9418
Residual		36.8304	0.6293	58.52	<.0001	0.05	35.6274	38.0956

The Level 2 covariance parameter estimates are listed by the row and column position in the unstructured covariance matrix for the random effects. The estimated **G** matrix is as follows:

$$\mathbf{G} = \begin{pmatrix} \sigma^2{}_{00} & \\ \sigma_{10} & \sigma^2{}_{11} \end{pmatrix} = \begin{pmatrix} \text{un}(1,1) & \\ \text{un}(2,1) & \text{un}(2,2) \end{pmatrix} = \begin{pmatrix} 4.8278 & \\ -0.1547 & 0.4127 \end{pmatrix}$$

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

Examining the estimates, the slope variance is statistically significant, which indicates that there are 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, which is consistent with the notion that the scores of impoverished students are raised to the level of affluent students? The correlation of -0.11 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 -0.74 to 0.43, which indicates that this relationship is not significant.

Solution for Fixed Effects					
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	12.6651	0.1898	159	66.72	<.0001
student_ses	2.3938	0.1181	7024	20.27	<.0001

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. From the significance of the variance component for this variable, it is evident that the SES effect varies significantly across schools.

End of Demonstration

Wrap-Up

Thank you for attending our SAS seminar.

Instructor email: Mike.Patetta@sas.com

Course links:

https://support.sas.com/edu/schedules.html?ctry=us&crs=BHLNM

31