Using SAS PROC MIXED to Fit Multilevel & Hierarchical Models

November 15, 2006

Charlie Hallahan

Overview

This presentation is based on the paper "**Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models**" by Judith Singer.

(*Journal of Educational and Behavioral Statistics, Winter 1998, Vol. 24, No. 4, pp. 323-355*)

The paper (and links to the data) can be found at http://gseweb.harvard.edu/~faculty/singer/.

This demo will cover the **Multilevel Model example** and next month will look at the **Growth Model example**.

Two-Level School Effects Models

The **school-effects models** are designed for data on individuals nested within naturally occurring hierarchies (e.g., students within classes, children within families, teachers within schools).

Multilevel models can be expressed in either of two ways:

1. **structural equations**: an equation is written at each level
2. **reduced form equation**: the structural equations are combined to form a single equation.

PROC MIXED requires the model to be expressed in reduced form.

Two-Level School Effects Models

The example to be used in the **High School and Beyond** data set used by **Bryk** and **Raudenbush** in their 1992 text on Multilevel Models.

The data set has data on **7,185 students** in **160 schools** (with anywhere from 14 to 67 students per school).

The variables to be used in this example are given below. The **Level 1 variables** are at the student level and the **Level 2 variables** are at the school level.

Level 1 variables:

mathach – math achievement score for each student. This will be the dependent variable in the models.

ses – student socio-economic status (centered at the grand mean)

Level 2 variables:

meanses – mean ses score for each school (centered at the grand mean)

sector – dummy variable for public (0) or private (1) school

Two-Level School Effects Models

Before estimating any models, it is worthwhile to illustrate how some simple **2-level models** can be first specified in **structural form** and then in **reduced form**.

No predictors : fixed intercept

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

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

Reduced-form model: $y_{ij} = \gamma_{00} + r_{ij}$ $\qquad r_{ij} \sim N\left(0, \sigma^2\right)$

```
PROC MIXED data = singer.hsb12;
       model mathach = /solution;
run;
```

PROC MIXED always assumes an intercept is in the model. This model just estimates the grand mean of the dependent variable `mathach`.

Two-Level School Effects Models

No predictors : random intercept

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

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

Reduced-form model: $y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \quad u_{0j} \sim N\left(0, \tau_{00}\right) \ , \ r_{ij} \sim N\left(0, \sigma^2\right)$

```
PROC MIXED data = singer.hsb12;
       class school;
       model mathach = /solution;
       random school;
run;
```

The **`random`** statement states that the variable ***`school`*** defines the hierarchy (or nesting or clustering) in the dataset and that this is to be treated as a **random effect**.

Two-Level School Effects Models

One predictor : fixed intercept and fixed slope

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

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

$\beta_{1j} = \gamma_{01}$

Reduced-form model: $y_{ij} = \gamma_{00} + \gamma_{01}x_{ij} + r_{ij}$ $\quad r_{ij} \sim N\left(0, \sigma^2\right)$

```
PROC MIXED data = singer.hsb12;
       model mathach = ses /solution;
run;
```

This is just a **fixed-parameter model** that could be estimated with **`PROC GLM`** or **`REG`**.

Two-Level School Effects Models

One predictor : random intercept and fixed slope

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

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

$\beta_{1j} = \gamma_{01}$

Reduced-form model: $y_{ij} = (\gamma_{00} + \gamma_{01} x_{ij}) + u_{0j} + r_{ij} \quad u_{0j} \sim N(0, \tau_{00}) \;,\; r_{ij} \sim N(0, \sigma^2)$

```
PROC MIXED data = singer.hsb12;
       class school;
       model mathach = ses/solution;
       random school;
run;
```

Two-Level School Effects Models

One predictor : random intercept and random slope

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

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

$\beta_{1j} = \gamma_{01} + u_{1j}$

Reduced-form model: $y_{ij} = (\gamma_{00} + \gamma_{01} x_{ij}) + (u_{0j} + u_{1j} x_{ij}) + r_{ij}$

$u_{0j} \sim N\left(0, \tau_{00}\right)$, $u_{1j} \sim N\left(0, \tau_{11}\right)$,

$\mathrm{Cov}(u_{0j}, u_{1j}) = \tau_{01}$, $r_{ij} \sim N\left(0, \sigma^2\right)$

```
PROC MIXED data = singer.hsb12;
       class school;
       model mathach = ses/solution;
       random school ses;
run;
```

Unconditional Means Model

Each of the previous models implies a particular structure for the error covariance matrix.

In particular, the "**No predictors: random intercept**" model (on page 6) is also called the ***Unconditional Means Model*** in **Singer** and is the first model estimated.

Recall that the Reduced-form Model is:

$$y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \quad u_{0j} \sim N\left(0, \tau_{00}\right) \;,\; r_{ij} \sim N\left(0, \sigma^2\right)$$

We assume that u_{0j} and r_{ij} are independent. The compound error term is

$$\xi_{ij} = u_{0j} + r_{ij} \qquad \text{(}\textit{error components}\text{)}$$

Unconditional Means Model

with variance

$$Var(\xi_{ij}) = Var(u_{0j} + r_{ij}) = Var(u_{0j}) + Var(r_{ij}) = \tau_{00} + \sigma^2 \quad (\textit{variance components}).$$

From the independence assumptions for u_{0j} and r_{ij}, it follows that

$$Cov(\xi_{ij}, \xi_{ij'}) = \tau_{00} \text{ for } j = j' \text{ and } = 0 \text{ for } j \neq j'.$$

Then $Cov(\xi)$ is **block - diagonal** with the size of the blocks depending on the number of observations within each cluster. For example, a block of size 3 would be:

$$\begin{bmatrix} \tau_{00} + \sigma^2 & \tau_{00} & \tau_{00} \\ \tau_{00} & \tau_{00} + \sigma^2 & \tau_{00} \\ \tau_{00} & \tau_{00} & \tau_{00} + \sigma^2 \end{bmatrix}$$

This variance-covariance structure is called ***compound symmetry***.

Unconditional Means Model

The PROC MIXED code can be written in two equivalent ways:

```
PROC MIXED data = singer.hsb12;
      class school;
      model mathach = /solution;
      random school;
run;

PROC MIXED data = singer.hsb12 noclprint covtest;
      class school;
      model mathach = /solution;
      random intercept / subject=school;
run;
```

In the 2nd way above, the option `noclprint` suppresses printing out the 160 values for the class variable `school`. The option `covtest` provides hypothesis tests for the variance and covariance terms. These tests require large samples in order to be considered valid.

Unconditional Means Model

Note that `PROC MIXED` always assumes the model contains an intercept term (the only fixed effect in this example).

A look at the 1st few observations in the dataset:

High School and Beyond data set

Obs	SCHOOL	MATHACH	SES	MEANSES	SECTOR
1	1224	5.876	-1.528	-0.428	0
2	1224	19.708	-0.588	-0.428	0
3	1224	20.349	-0.528	-0.428	0
4	1224	8.781	-0.668	-0.428	0
5	1224	17.898	-0.158	-0.428	0
6	1224	4.583	0.022	-0.428	0
7	1224	-2.832	-0.618	-0.428	0
8	1224	0.523	-0.998	-0.428	0
9	1224	1.527	-0.888	-0.428	0
10	1224	21.521	-0.458	-0.428	0

Unconditional Means Model

Model Information

Data Set	SINGER.HSB12
Dependent Variable	MATHACH
Covariance Structure	Variance Components
Subject Effect	SCHOOL
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Default covariance structure is `Variance Components` (resulting in *compound symmetry* in this case).

Default estimation method is `REML`.

Dimensions

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

In this example we have 2 Covariance Parameters (σ^2 and τ_{00}). The matrices X and Z have 1 column each for the intercept.

Number of Observations

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

Unconditional Means Model

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.

Covariance Parameter Estimates

	Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
τ_{00}	Intercept	SCHOOL	8.6097	1.0778	7.99	<.0001
σ^2	Residual		39.1487	0.6607	59.26	<.0001

Fit Statistics

-2 Res Log Likelihood	47116.8
AIC (smaller is better)	47120.8
AICC (smaller is better)	47120.8
BIC (smaller is better)	47126.9

Unconditional Means Model

Solution for Fixed Effects

	Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
γ_{00}	Intercept	12.6370	0.2443	159	51.72	<.0001

Covariance parameter estimates:

$\hat{\tau}_{00} = 8.6097$ and $\hat{\sigma}^2 = 39.1487$, so the estimate of the total variance is $\hat{\xi}^2 = \hat{\tau}_{00} + \hat{\sigma}^2 = 47.7584$.

Recall that $\hat{\tau}_{00}$ represents the variability of *Mathach* within a given school and $\hat{\sigma}^2$ represents the variability of an individual student's score. Also, for two students within the same school, $\text{cov}(\xi_{ij}, \xi_{i'j}) = \tau_{00}$. The quantity $\rho = \dfrac{\tau_{00}}{\tau_{00} + \sigma^2}$, which is a measure of the amount of clustering within each school, is called the **intraclass correlation**.

In this example, $\hat{\rho}$=0.18. With this degree of correlation among the residuals, OLS would not be appropriate.

Unconditional Means Model

When comparing multiple models with the same fixed effects but different random effects, the two goodness-of-fit criteria **AIC** and **SBC** are commonly used.

When doing such comparisons, the estimation method **ML** should be used instead of the default **REML**.

Smaller (in absolute terms) is better for these criteria.

The idea now is to fit a series of different models for this example.

Including Effects of School-Level (level-2) Predictors

The **unconditional means model** provides a **baseline for comparison** with more complex models.

We now add the school level (level 2) variable, **MEANSES**, the average SES score for each school.

Recall that this variable has been **centered at its mean** (making interpretation of the intercept straightforward).

A **conditional model**:

Structural form: $Y_{ij} = \beta_{0j} + r_{ij}$ and $\beta_{0j} = \gamma_{00} + \gamma_{01} MEANSES_j + u_{0j}$

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

Reduced form : $Y_{ij} = \left[\gamma_{00} + \gamma_{01} MEANSES_j\right] + \left[u_{0j} + r_{ij}\right] = \textit{fixed part} + \textit{random part}$

Including Effects of School-Level (level-2) Predictors

```
title2 "Conditional Means Model with MEANSES";
proc mixed data=singer.hsb12 noclprint covtest;
      class school;
      model mathach = meanses/solution ddfm=bw;
      random intercept / subject=school;
run;
```

The only changes from the unconditional mean model is adding the fixed effect `meanses` to the `model` statement and requesting that the denominator degrees of freedom (`ddfm=bw`) be calculated using the ***"between/within"*** method.

This option affects the F-tests for fixed effects. The default choice results in `df = 7025` (representing all the observations) while this option give `df = 158` (representing the number of schools in the sample).

Including Effects of School-Level (level-2) Predictors

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	47201.23573408	
1	2	46961.28490236	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
Intercept	SCHOOL	2.6357	0.4036	6.53	<.0001
Residual		39.1578	0.6608	59.26	<.0001

Fit Statistics

-2 Res Log Likelihood	46961.3
AIC (smaller is better)	46965.3
AICC (smaller is better)	46965.3
BIC (smaller is better)	46971.4

Including Effects of School-Level (level-2) Predictors

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.6495	0.1492	158	84.77	<.0001
MEANSES	5.8635	0.3613	158	16.23	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
MEANSES	1	158	263.37	<.0001

Including Effects of School-Level (level-2) Predictors

Interpretation of output :

Fixed Effects :

INTERCEPT: $\hat{\gamma}_{00} = 12.65$ estimates the mean math achievement when the remaining predictors (MEANSES) are 0. Since MEANSES is centered at its grand mean, then the prediction of math achievement for a school with average MEANSES is 12.65.

MEANSES: $\hat{\gamma}_{01} = 5.86$ says that for each increase of 1 unit above the mean SES score the mean math achievement score increases by 5.86 units. With a *t - statistic* of 16.22, this effect is significant.

Including Effects of School-Level (level-2) Predictors

Interpretation of output :

Random Effects :

$\hat{\sigma}^2 = 39.16$ estimates the residual variance. Note that the estimate has hardly changed from the **Unconditional Mean Model** estimate of 39.15.

$\hat{\gamma}_{00} = 2.65$ estimates the variation between schools. Note that the estimate has decreased significantly from the **Unconditional Mean Model** estimate of 8.61. We are now estimating a **conditional model** and this large decrease in this variance component shows that adding the fixed effect MEANSES accounts for a large proportion of the school-to-school variation.

Since (8.61 - 2.65)/8.61 = .69, we could say that 69% of the explainable variation in school mean math achievement scores is explained by MEANSES.

Including Effects of School-Level (level-2) Predictors

With an estimated residual variance of *39.16* (and a *z - statistic* of *6.53* indicating significance), the question can be raised if any additional explainable variation (besides the fixed effect **MEANSES**) can be found.

residual intraclass correlation : represents the intraclass correlation of schools "*of comparable SES*".

In this case it is calculated as $2.63/(2.63 + 39.16) = .06$.

It can be interpreted as a **partial correlation** and measures the similarity in math achievement among students within schools *after controlling* for the effect of **MEANSES**.

Including Effects of Student-Level (level-1) Predictors

In searching for more explainable measures of the variation in mean math achievement scores, we could add ***student-level (level-1)*** variables to the model.

We'll do this one step at a time and initially consider a model with only a level-1 explanatory variable, namely, **SES**.

A simple random coefficient model:

$$Y_{ij} = \beta_{0j} + \beta_{1j} SES_{ij} + r_{ij}, \quad \beta_{0j} = \gamma_{00} + u_{0j}, \quad \beta_{1j} = \gamma_{01} + u_{1j},$$

$$\text{where } r_{ij} \sim N\left(0, \sigma^2\right) \quad \text{and} \quad \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right]$$

reduced form: $Y_{ij} = \left(\gamma_{00} + \gamma_{01} SES_{ij}\right) + \left(u_{0j} + u_{1j} SES_{ij} + r_{ij}\right)$

Including Effects of Student-Level (level-1) Predictors

There are several differences between this model and the **Unconditional Means Model.**

1. a single *level-1 predictor*, **SES**. is included

2. **SES** has a *random coefficient* which is reflected in the random effects part of the model, i.e., the effect of a student's SES score can vary randomly across schools.

3. There are now **three variance components**, two variances and a covariance between the *intercept-random effect* and the *SE-random effect*.

Including Effects of Student-Level (level-1) Predictors

Even though the above model could easily be estimated with **PROC MIXED**, the author chooses not to do so because of a problem with the interpretation of the parameters as the model is currently specified.

When $SES_{ij} = 0$, then β_{oj} represents the mean math achievement score for a student with an average **SES** score **over all students** (since SES_{ij} is centered on its grand mean).

We would like β_{oj} to represent the mean math achievement score for a student ***in school j*** whose **SES** score is equal to the average level for that school.

This can be accomplished by centering SES_{ij} on $MEANSES_j$, i.e., create the new variable $CSES_{ij} = SES_{ij} - MEANSES_j$.

Now when $CSES_{ij} = 0$, we are talking about an observation where the student's **SES** score is equal to his/her school mean **SES** score.

Including Effects of Student-Level (level-1) Predictors

A re-specified random coefficient model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(SES_{ij} - S\bar{E}S_{j}) + r_{ij}, \quad \beta_{0j} = \gamma_{00} + u_{0j}, \quad \beta_{1j} = \gamma_{10} + u_{1j},$$

$$\text{where } r_{ij} \sim N\left(0, \sigma^2\right) \quad \text{and} \quad \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right]$$

reduced form: $Y_{ij} = \left(\gamma_{00} + \gamma_{10}(SES_{ij} - S\bar{E}S_{j})\right) + \left(u_{0j} + u_{1j}(SES_{ij} - S\bar{E}S_{j}) + r_{ij}\right)$

Including Effects of Student-Level (level-1) Predictors

```
* create SES variable centered around the school level SES;
data hsb12;
        set singer.hsb12;
        cses = ses - meanses;
run;

title2 "Conditional Means Model with CSES";
proc mixed data=hsb12 noclprint covtest noitprint;
        class school;
        model mathach= cses/solution ddfm=bw notest;
        random intercept cses/ subject=school type=un;
run;
```

Note that the `random` statement now specifies a random intercept and a random coefficient for *cses*.

The option `type=un` specifies an unstructured covariance matrix.

This option is common in **school effects models** (according to the author).

A different covariance structure is used for the **individual growth models**.

Including Effects of Student-Level (level-1) Predictors

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOL	8.6769	1.0786	8.04	<.0001
UN(2,1)	SCHOOL	0.05075	0.4062	0.12	0.9006
UN(2,2)	SCHOOL	0.6940	0.2808	2.47	0.0067
Residual		36.7006	0.6258	58.65	<.0001

Fit Statistics

-2 Res Log Likelihood	46714.2
AIC (smaller is better)	46722.2
AICC (smaller is better)	46722.2
BIC (smaller is better)	46734.5

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
3	1065.70	<.0001

Including Effects of Student-Level (level-1) Predictors

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.6493	0.2445	159	51.75	<.0001
cses	2.1932	0.1283	7024	17.10	<.0001

Interpreting the output from models with level-1 predictors

Focus first on the fixed effects:

$\hat{\gamma}_{00} = 12.65$ says that the estimated average school mean math achievement score, controlling for student SES, is 12.65

$\hat{\gamma}_{01} = 2.19$ says that the estimated average slope representing the relationship between student SES and math achievement is 2.19.

The t - *statistics* indicate that both estimates are highly significant.

Including Effects of Student-Level (level-1) Predictors

Covariance parameter estimates:

$$\begin{pmatrix} \hat{\tau}_{00} & \hat{\tau}_{01} \\ \hat{\tau}_{01} & \hat{\tau}_{11} \end{pmatrix} = \begin{pmatrix} 8.68 & 0.05 \\ 0.05 & 0.69 \end{pmatrix}$$

$\hat{\tau}_{00} = 8.68$ represents the variability of the random intercepts.

$\hat{\tau}_{01} = 0.05$ represents the covariance between the random intercepts and slopes.

$\hat{\tau}_{11} = 0.69$ represents the variability of the random slopes.

Conclusions:

1. The intercepts are very variable. i.e., schools differ in average math achievement scores even after controlling for the effects of student SES.
2. The slopes also have a non-zero variance ($p = 0.0135$).
3. There is little correlation between the intercepts & slopes ($\hat{\tau}_{01} = 0.05, p = .9006$). i.e., there is no evidence that the effects of student SES on math achievement differs depending upon the average math achievement in the school.

Including Effects of Student-Level (level-1) Predictors

We can ask the question: How much of the within school variance in math achievement is explained by student SES?

One way to address this question is to compare the estimates of the residual variance, σ^2, between the unconditional model and the model conditioned on SES.

(39.15 - 36.70)/39.15 = 0.06 or 6%. So inclusion of student level SES explains 6% of the explainable variation within schools.

Recall that when adding the variable MEANSES to the unconditional model, the variation between school-level math achievement scores fell from 8.61 to 2.64.

Including Effects of Student-Level (level-1) Predictors

We saw before that the inclusion of **school-level SES** reduced the **variation in school-level math achievement scores** from 8.61 to 2.64.

Thus, comparatively speaking, **school-level SES** explains more of the variation in **school-level math achievement** scores than **student-level SES** explains the variation in **student-level math achievement scores**.

Discussions of "explained variation" in multilevel models can be misleading. The author recommends **Snijders & Bosker (1994)** for a discussion of this issue.

Including Both Level-1 and Level-2 Predictors

The next model will contain one level-1 variable (student-level **SES**) and two level-2 variables (school-level SES, **MEANSES**, and a dummy variable indicating whether a school is public, 0, or private, 1, **SECTOR**).

The author strongly recommends writing out the model in mathematical notation before writing any SAS code.

The model is first written in **structural form**, and then the **reduced form** is derived.

Including Both Level-1 and Level-2 Predictors

Structural model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(SES_{ij} - S\overline{E}S_j) + r_{ij},$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01} MEANSES_j + \gamma_{02} SECTOR_j + u_{0j},$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} MEANSES_j + \gamma_{12} SECTOR_j + u_{1j},$$

where $r_{ij} \sim N\left(0, \sigma^2\right)$ and $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right]$

Reduced form: $$\begin{aligned} Y_{ij} &= [\gamma_{00} + \gamma_{01} MEANSES_j + \gamma_{02} SECTOR_j + \gamma_{10}(SES_{ij} - S\overline{E}S_j) \\ &\quad + \gamma_{11} MEANSES_j (SES_{ij} - S\overline{E}S_j) + \gamma_{12} SECTOR_j (SES_{ij} - S\overline{E}S_j)] \\ &\quad + \left(u_{0j} + u_{1j}(SES_{ij} - S\overline{E}S_j) + r_{ij}\right) \\ &= \text{"fixed effects"} + \text{"random effects"} \end{aligned}$$

Including Both Level-1 and Level-2 Predictors

```
title2 "Model with both level-1 and level-2 explanatory variables";
proc mixed data=hsb12 noclprint covtest noitprint;
        class school;
        model mathach = meanses sector cses meanses*cses sector*cses
                      / solution ddfm=bw notest;
        random intercept cses / type=un sub=school;
run;
```

Note that ***interactions*** can be represented by simply including an asterisk (*) between the relevant variables.

Including Both Level-1 and Level-2 Predictors

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOL	2.3817	0.3717	6.41	<.0001
UN(2,1)	SCHOOL	0.1926	0.2045	0.94	0.3464
UN(2,2)	SCHOOL	0.1014	0.2138	0.47	0.3177
Residual		36.7212	0.6261	58.65	<.0001

Fit Statistics

-2 Res Log Likelihood	46503.7
AIC (smaller is better)	46511.7
AICC (smaller is better)	46511.7
BIC (smaller is better)	46524.0

Including Both Level-1 and Level-2 Predictors

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.1136	0.1988	157	60.93	<.0001
MEANSES	5.3391	0.3693	157	14.46	<.0001
SECTOR	1.2167	0.3064	157	3.97	0.0001
cses	2.9388	0.1551	7022	18.95	<.0001
MEANSES*cses	1.0389	0.2989	7022	3.48	0.0005
SECTOR*cses	-1.6426	0.2398	7022	-6.85	<.0001

Interpreting the output: fixed effects

fixed effects: all are significant

writing the separate models represented by the dummy variable SECTOR:

Public: MATHACH = 12.11 + 5.34 MEANSES + 2.94 CSES + 1.03 MEANSES*CSES

Catholic: MATHACH = 13.33 + 5.34 MEANSES + 1.30 CSES + 1.03 MEANSES*CSES

Including Both Level-1 and Level-2 Predictors

The significant interaction between **MEANSES** and **CSES** says that the slope for the **student-level SES** differs according to the **school-level SES** and the significant interaction between **SECTOR** and **CSES** says that the effect of **student-level SES** on **MATHACH** differs for Public and Catholic schools.

Because **MEANSES** has a grand mean of zero and **CSES** is centered at its school mean, the intercept has a straightforward interpretation:

> The average Public school math achievement score is 12.11 and the average Catholic school score is 13.33 and these means are significantly different.
>
> Student and school **SES** are associated with math achievement in both sectors, although the magnitude of the student effect differs across sectors.
>
> There is also an interaction between student and school **SES**. In both sectors, the slope for student **SES** is higher in schools with higher mean **SES** levels.

Including Both Level-1 and Level-2 Predictors

Interpreting the output: random effects

The **variance component for intercepts** (τ_{00}) is significant - suggesting that there is additional variation in school mean achievement levels not accounted for by these three factors and their interactions.

The **variance component for slopes** (τ_{11}) is not significant, nor is the covariance (τ_{01}) between the random intercepts and slopes.

A simpler model with **fixed slopes** and **random intercept** could then be tried.

Including Both Level-1 and Level-2 Predictors

```
title2 "Model with both level-1 and level-2 explanatory variables";
title3 "Dropping the random slopes";
proc mixed data=hsb12 noclprint covtest noitprint;
        class school;
        model mathach = meanses sector cses meanses*cses sector*cses
                      / solution ddfm=bw notest;
        random intercept / sub=school;
run;
```

Including Both Level-1 and Level-2 Predictors

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	SCHOOL	2.3752	0.3709	6.40	<.0001
Residual		36.7661	0.6207	59.24	<.0001

Fit Statistics

-2 Res Log Likelihood	46504.8
AIC (smaller is better)	46508.8
AICC (smaller is better)	46508.8
BIC (smaller is better)	46514.9

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	12.1138	0.1986	157	60.98	<.0001
MEANSES	5.3429	0.3690	157	14.48	<.0001
SECTOR	1.2146	0.3061	157	3.97	0.0001
cses	2.9358	0.1507	7022	19.48	<.0001
MEANSES*cses	1.0441	0.2910	7022	3.59	0.0003
SECTOR*cses	-1.6421	0.2331	7022	-7.04	<.0001

Including Both Level-1 and Level-2 Predictors

Comparing the two models using the fit criteria

	AIC	SBC	-2LL
random intercept and slopes	-23.255.8	-23.269.6	46.503.67
random intercepts	-23.254.4	-23.261.3	46.504.79

Based on the AIC and SBC, the random intercepts model is preferred.

The difference in -2LL is only 1.12, which is distributed as χ^2 with 2 degrees of freedom.