SOCY7708: Hierarchical Linear Modeling
Instructor: Natasha Sarkisian

Two-level HLM Models

We'll be working with "High School and Beyond" data -- I also placed these data files on the course website as Stata data files: hsb1.dta and hsb2.dta – one file for each level, and hsb.dta – the two levels combined.

The level-1 file (hsb1.dta) has 7185 cases (students) and 5 variables:
id – school number
minority – an indicator of students' ethnicity (1=minority, 0=other)
female – an indicator of students' gender (1=female, 0=male)
ses – a standardized scale constructed from measures of parental occupation, education, and income
mathach – a measure of mathematics achievement

The level-2 file (hsb2.dta) has 160 cases (schools) and 7 variables:
id – school number
size – school enrollment
sector – 1=Catholic, 0=public
pracad – proportion of students in the academic track
disclim – a scale measuring disciplinary climate
himnty – 1=more than 40% minority enrollment, 0=less than 40%
meanses – mean of the SES values for the students in each school (generated as group means from level 1 file).

It is essential that both files contain group identifier (in this case, school ID).

The combined file (hsb.dta) has 7185 cases (students) and 11 variables – same as above. Note that school-level variables now have 7185 observations, but they are the same for all students within each school. This is called the disaggregation of level 2 predictors. Let's examine these files.

```
. use hsb1.dta, clear

. sum

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
          id |          0
    minority |      7,185     .274739    .4464137          0          1
      female |      7,185    .5281837    .4992398          0          1
         ses |      7,185    .0001434    .7793552     -3.758      2.692
     mathach |      7,185    12.74785    6.878246     -2.832     24.993

. use hsb2.dta, clear

. sum

    Variable |        Obs        Mean    Std. dev.       Min        Max
```

```
-------------+---------------------------------------------------------
          id |          0
        size |        160    1097.825    629.5064        100       2713
      sector |        160       .4375    .4976359          0          1
      pracad |        160    .5139375    .2558967          0          1
     disclim |        160    -.015125    .9769777     -2.416      2.756
-------------+---------------------------------------------------------
     himinty |        160        .275    .4479162          0          1
     meanses |        160   -.0001875    .4139731     -1.188       .831


. use hsb.dta, clear

. sum

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
          id |          0
    minority |      7,185     .274739    .4464137          0          1
      female |      7,185    .5281837    .4992398          0          1
         ses |      7,185    .0001434    .7793552     -3.758      2.692
     mathach |      7,185    12.74785    6.878246     -2.832     24.993
-------------+---------------------------------------------------------
        size |      7,185    1056.862    604.1725        100       2713
      sector |      7,185    .4931106    .4999873          0          1
      pracad |      7,185    .5344871    .2511861          0          1
     disclim |      7,185   -.1318694    .9439882     -2.416      2.756
     himinty |      7,185    .2800278    .4490438          0          1
-------------+---------------------------------------------------------
     meanses |      7,185    .0061385    .4135539     -1.188       .831
```

If we are interested in providing descriptive statistics for our level 2 variables, we should either use level 2 dataset, or calculate summary stats in the combined dataset for only 1 observation per upper level unit, e.g.:

```
. egen tag=tag(id)

. tab tag

    tag(id) |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      7,025       97.77       97.77
          1 |        160        2.23      100.00
------------+-----------------------------------
      Total |      7,185      100.00

. sum size- meanses if tag==1

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
        size |        160    1097.825    629.5064        100       2713
      sector |        160       .4375    .4976359          0          1
      pracad |        160    .5139375    .2558967          0          1
     disclim |        160    -.015125    .9769777     -2.416      2.756
     himinty |        160        .275    .4479162          0          1
-------------+---------------------------------------------------------
     meanses |        160   -.0001875    .4139731     -1.188       .831
```

Next, let's try to estimate the simplest possible model -- it is known as the fully unconditional model (FUM), where no predictors are specified at either level 1 or level 2. Our dependent variable will be mathach.

Model 0. Unconditional model with random intercept (a.k.a. intercept-only model, or one way ANOVA with random intercept):

LEVEL 1 MODEL

$$\text{MATHACH}_{ij} = \beta_{0j} + r_{ij}$$

LEVEL 2 MODEL

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

MIXED MODEL

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

γ_{00} is the grand mean (i.e. average intercept) – this is the fixed component of the model (fixed effect)

The two random components are:
$r_{ij} \sim N(0, \sigma^2)$
$u_{0j} \sim N(0, \tau_{00})$

Note that this model is very similar to a one-way ANOVA model utilizing the grouping variable as a nominal-level variable. What distinguishes the two is the random variable nature of u_{0j} -- in a regular one-way ANOVA, each u_{0j} is a fixed number, in a sense it is the value of a dummy-variable indicator for that specific group. In HLM models, however, u_{0j} is modeled as a random variable rather than a set of fixed coefficients.

Estimating the fully unconditional model is useful as a preliminary step in a hierarchical data analysis. Its most important function is to provide the information about outcome variability at each of the two levels. Sigma (σ) will provide the information about level-1 (within-group) variability, and tau (τ) will provide the information on level-2 (between-group) variability. Running this model allows us to decompose the variance in the dependent variable into variance components for each hierarchical level -- into within-group and between-group variance. This model does not explain anything, but it allows us to evaluate whether there is variation across groups, and how much of it. That's why it is always a good idea to run this basic model when starting the analyses – it's the null model of our regression analysis. If we find that there is no significant between-group variation, then there is no need for a hierarchical model.

The proportion of variance due to group-level variation in means can be calculated as
$\rho = \tau_{00} / (\sigma^2 + \tau_{00})$
and it represents the *intra-class correlation coefficient*. It can be interpreted as the proportion of variance explained by the grouping structure in the population.

<u>Running the model:</u>

```
. mixed mathach || id:

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160
                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67
                                                Wald chi2(0)      =          .
Log likelihood = -23557.905                     Prob > chi2       =          .
------------------------------------------------------------------------------
     mathach | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
       _cons |   12.63707   .2436178    51.87   0.000     12.15959    13.11455
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                  var(_cons) |    8.55352   1.068642       6.69575    10.92674
-----------------------------+------------------------------------------------
               var(Residual) |   39.14839   .6606469      37.87473    40.46489
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 983.92        Prob >= chibar2 = 0.0000
```

Only one fixed effect is estimated in this model – that's the average value of the outcome across all individuals – here, the average math achievement is estimated to be 12.64.

The main thing we have to conclude from examining this output is that there is a substantial amount of school-level variation in math achievement. The intra-class correlation is:
$\rho = \tau_{00} / (\tau_{00} + \sigma^2) =$ 8.55 /(8.55+39.15)= .18
So 18% of the total variance is on the school level.

Or we could calculate that using the quantities stored after the model:

```
. ereturn list

scalars:
               e(rank) =  3
                e(k_f) =  1
                e(k_r) =  2
                  e(k) =  3
               e(k_rs) =  2
               e(k_rc) =  0
              e(k_res) =  0
          e(converged) =  1
                 e(ic) =  1
                 e(ll) =  -23557.90510815813
               e(ll_c) =  -24049.86601856097
               e(df_c) =  1
             e(chi2_c) =  983.9218208056773
                e(p_c) =  2.8059555978e-216
                  e(N) =  7185
           e(nrgroups) =  1
              e(small) =  0
               e(df_m) =  0
                  e(p) =  .
               e(chi2) =  .
                 e(rc) =  0
```

```
macros:
             e(cmdline) : "mixed mathach || id:"
  e(datasignaturevars) : "mathach id"
      e(datasignature) : "7185:2:1423354169:3944075123"
           e(chi2type) : "Wald"
         e(rstructlab) : "Independent"
         e(rstructure) : "independent"
          e(estat_cmd) : "mixed_estat"
            e(predict) : "mixed_p"
              e(redim) : "1"
              e(ivars) : "id"
           e(vartypes) : "Identity"
             e(revars) : "_cons"
                e(cmd) : "mixed"
              e(title) : "Mixed-effects ML regression"
          e(technique) : "nr"
          e(ml_method) : "d0"
                e(opt) : "moptimize"
             e(method) : "ML"
          e(optmetric) : "matsqrt"
             e(depvar) : "mathach"
         e(properties) : "b V"

matrices:
                  e(b) :  1 x 3
                  e(V) :  3 x 3
              e(g_max) :  1 x 1
              e(g_avg) :  1 x 1
              e(g_min) :  1 x 1
                e(N_g) :  1 x 1

functions:
             e(sample)

. mat list e(b)

e(b)[1,3]
      mathach:  lns1_1_1:   lnsig_e:
        _cons       _cons       _cons
y1   12.63707  1.0731714  1.8336797

. di exp(e(b)[1,2])^2
8.5535197

. di exp(e(b)[1,2])^2/(exp(e(b)[1,2])^2 + exp(e(b)[1,3])^2)
.17931188
```

Even easier, however, is to use estat icc command:

```
. estat icc

Intraclass correlation
------------------------------------------------------------------------------
                       Level |        ICC   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
                          id |   .1793119   .0185938       .145709    .2186805
------------------------------------------------------------------------------
```

If we would like to test whether that variance component is statistically significant, we could look at the last line, LR test vs. linear model. The p-value there is very small, which means this model is significantly different from regular OLS and therefore level 2 variance is significant.

After estimating a null model and assuring that we observe a significant amount of group-level variance, we proceed to build a multilevel explanatory model. A typical approach is to build such a model from bottom up.

Model 1. Conditional model with random intercept (one way ANCOVA with random intercept)

LEVEL 1 MODEL

$$MATHACH_{ij} = \beta_{0j} + \beta_{1j}(SES_{ij}) + r_{ij}$$

LEVEL 2 MODEL

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

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

MIXED MODEL

$$MATHACH_{ij} = \gamma_{00} + \gamma_{10} * SES_{ij} + u_{0j} + r_{ij}$$

```
. mixed mathach ses || id:

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160
                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67
                                                Wald chi2(1)      =     511.98
Log likelihood = -23320.502                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
     mathach | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         ses |   2.391499   .1056926    22.63   0.000     2.184346    2.598653
       _cons |   12.65762   .1873212    67.57   0.000     12.29048    13.02477
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                  var(_cons) |   4.728519   .6486766      3.613716     6.18723
-----------------------------+------------------------------------------------
               var(Residual) |   37.02979    .625279      35.82432    38.27582
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 456.94        Prob >= chibar2 = 0.0000
```

Note that we now estimate two fixed effects – the intercept and the effect of student's SES. The intercept γ_{00} is no longer the average math achievement – it is now math achievement for someone with all predictors equal to zero. In this case, it's math achievement for someone with SES=0, but because the SES scale was designed to have a mean of 0, the intercept (12.66) is essentially the math achievement for someone with average SES. The effect of SES, γ_{10}, can be interpreted as follows: one unit increase in SES is associated with 2.39 unit increase in one's math achievement. So math achievement for someone with SES being 1 unit above the mean would be:
12.66+2.39=15.05

Note that each β_{0j} is now the mean outcome for each group (i.e. school) adjusted for the differences among these groups in SES.

As we now accounted for some portion of the variance by controlling for SES, we can calculate the adjusted intra-class correlation: ρ=4.73/(4.73+37.03)= .11

```
. estat icc

Residual intraclass correlation
------------------------------------------------------------------------------
                       Level |        ICC   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
                          id |   .1132354   .0139342      .0886613    .1435479
------------------------------------------------------------------------------
```

The decrease in ρ from .18 to .11 reflects a reduction in the relative share of between-school variance when we control for student SES. But there is still significant variation across schools.

We could also calculate the proportion of variance explained at each level by comparing the current variance estimates to those in the null model. (This is the easiest method recommended by Bryk and Raudenbush; another method is suggested by Snijders and Bosker and described in the Multilevel Modeling book by Douglas Luke, published by Sage, p.35-37):

(8.55 - 4.73)/8.55 = .45
(39.15 - 37.03)/ 39.15 = .05

So controlling for individuals' SES explained 45% of between-school variance, and 5% of within-school variance in math achievement. We could also calculate the total percentage of variance explained:
(39.15+8.55-4.73-37.03)/(39.15+8.55)= .12
So students' SES explained 12% of the total variance in math achievement.

Or we could calculate that using values that are stored in matrices:

```
. qui mixed mathach || id:

. mat base=e(b)

. qui mixed mathach ses || id:

. mat ses=e(b)

. mat list base

base[1,3]
      mathach:  lns1_1_1:   lnsig_e:
        _cons       _cons       _cons
y1   12.63707  1.0731714  1.8336797
. mat list ses

ses[1,4]
      mathach:   mathach:  lns1_1_1:   lnsig_e:
          ses       _cons       _cons       _cons
y1  2.3914994  12.657623  .77680602  1.8058613
```

```
. di exp(base[1,2])^2
8.5535197

. di (exp(base[1,2])^2 - exp(ses[1,3])^2)/exp(base[1,2])^2
.44718443

. di (exp(base[1,3])^2 - exp(ses[1,4])^2)/exp(base[1,3])^2
.05411731

. di (exp(base[1,2])^2+ exp(base[1,3])^2 - exp(ses[1,3])^2 -
exp(ses[1,4])^2)/(exp(base[1,2])^2 + exp(base[1,3])^2)
.12459891
```

Let's take this one step further. So far we assumed that an individual student's SES would have the same impact on his or her math achievement regardless of the school where that student is studying. Let's relax that assumption.

Model 2. Model with random intercept and random slopes (one way ANCOVA with random intercept and slopes)

LEVEL 1 MODEL

$$\text{MATHACH}_{ij} = \beta_{0j} + \beta_{1j}(\text{SES}_{ij}) + r_{ij}$$

LEVEL 2 MODEL

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

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

Here, level-1 slopes are allowed to vary across level-2 units. But we do not try to predict that variation – only describe it.

Now we have:
γ_{00} is the average intercept across the level-2 units (grand mean of math achievement controlling for SES – i.e. the mean for someone with SES=0)
γ_{10} is the average SES slope across the level-2 units (i.e. average effect of SES across schools)
u_{0j} is the unique addition to the intercept associated with level-2 unit j (indicates how the intercept for school j differs from the grand mean)
u_{1j} is the unique addition to the slope associated with level-2 unit j (indicates how the effect of SES in school j differs from the average effect of SES for all schools)

Note that:

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left(0, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right)$$

Our tau matrix now contains the variance in the level-1 intercepts (τ_{00}), the variance in level-1 slopes (τ_{11}), as well as the covariance between level-1 intercepts and slopes ($\tau_{01} = \tau_{10}$). We will specify the covariance(unstructured) option in our mixed command – that is because we want to allow random effects to correlate with each other; if we do not, that would be too restrictive since usually random effects for intercepts and slopes are correlated.

```
. mixed mathach ses || id: ses, cov(unstructured)

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160
                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67
                                                Wald chi2(1)      =     414.12
Log likelihood = -23318.235                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     mathach | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         ses |   2.394934   .1176881    20.35   0.000      2.16427    2.625598
       _cons |   12.66559   .1890996    66.98   0.000     12.29496    13.03621
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                    var(ses) |   .3983418   .2324055      .1269513    1.249898
                  var(_cons) |   4.785235   .6648578      3.644499    6.283024
              cov(ses,_cons) |  -.1558654   .2956225     -.7352749    .4235441
-----------------------------+------------------------------------------------
               var(Residual) |   36.83154   .6293445      35.61847    38.08592
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 461.47                Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

Here, almost like in the previous model, the math achievement for someone with average SES (SES=0) is 12.65; each unit increase in SES is associated with 2.4 units increase in math achievement. But, examining variance components, we notice that there is some variation in slopes – that variance is .4. So now we allow for SES effects to vary across schools; 2.4 is the effect for an average school.

Here, if we want to divide the unexplained variance into within-school and between-school, we need to take into account the covariance: level 1 component is simply 36.83, but level 2 component is (4.79+0.4+2*-0.16)= 4.87. But ultimately, to calculate variance explained, it is better to reestimate the model without any random slopes and use level 1 and level 2 variance from that model – that would be our previous model, since we didn't add any additional explanatory variables, our variance explained is exactly the same as in that previous model.

Note that the covariance value indicates how much intercepts and slopes covary: in our example, there is a negative correlation between intercepts and slopes. That is, the higher the intercept, the smaller the slope (i.e. if the school level of math achievement is high, the effect of SES in that school is smaller). We can see this as a variance-covariance matrix:

```
. estat recov

Random-effects covariance matrix for level id

             |       ses      _cons
-------------+----------------------
         ses |  .3983418
       _cons | -.1558654   4.785235
```

Or expressed as correlations (here, the strength is easier to interpret because correlations are standardized and range between -1 and 1):

```
. estat recov, corr

Random-effects correlation matrix for level id

             |       ses      _cons
-------------+----------------------
         ses |         1
       _cons | -.1128938          1
```

We can test whether there is significant variance in SES slopes:

```
. qui mathach ses || id: ses, cov(unstructured)

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df        AIC        BIC
-------------+---------------------------------------------------------------
           . |      7,185          .  -23318.23       6   46648.47   46689.75
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.

. est store slope

. qui mixed mathach ses || id:

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df        AIC        BIC
-------------+---------------------------------------------------------------
           . |      7,185          .   -23320.5       4      46649   46676.52
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.
```

The BIC value for the model with SES slope variance is higher than for the model without SES slope is actually larger (46689.75 vs. 46676.52), and the difference is over 10, so there is strong evidence in favor of the model without the random slope of SES. Let's conduct a likelihood ratio test:

```
. lrtest . slope

Likelihood-ratio test
Assumption: . nested within slope

 LR chi2(2) =   4.54
Prob > chi2 = 0.1035

Note: The reported degrees of freedom assumes the null hypothesis is not on the
boundary of the parameter space. If this is not true, then the reported test is
conservative.
```

This confirms that jointly, SES slope variance and slope/intercept covariance are not statistically significant. But for now, we will continue exploring models.

Model 3. Means-as-outcomes model (a.k.a. Intercepts as outcomes)

LEVEL 1 MODEL

$$\text{MATHACH}_{ij} = \beta_{0j} + r_{ij}$$

LEVEL 2 MODEL

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

This model allows us to predict variation in the levels of math achievement using level-2 variables. If we would attempt to do this using regular OLS, we would be artificially inflating the sample size and pretend we have 7185 data points to evaluate the effect of type of school (Catholic vs public), when in fact it's only 160 schools. Aggregating the data to school level would be more acceptable, but we would not have any assessment of within-school variation. Note, however, that the sample size for level 2 becomes important as soon as you try to include predictors at this level! You should try to have at least 10 level 2 cases per each level 2 variable you use.

```
. mixed mathach sector || id:

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160
                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67
                                                Wald chi2(1)      =      41.34
Log likelihood = -23539.553                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     mathach | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      sector |   2.804807   .4362268     6.43   0.000     1.949818    3.659796
       _cons |   11.39306   .2909743    39.15   0.000     10.82276    11.96336
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                  var(_cons) |   6.579583   .8490659       5.10922    8.473096
-----------------------------+------------------------------------------------
               var(Residual) |   39.15165   .6607522      37.87779    40.46836
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 715.25        Prob >= chibar2 = 0.0000
```

Here, we see a positive effect of Catholic schools on math achievement – the average achievement of Catholic schools is 2.8 units higher than for public schools. The intercept now is an average value for a public school student. There is, nevertheless, significant school-level variance remaining. As we did with earlier models, we can calculate the percentage of variance in math achievement explained by school type. Note that here we only explain level 2 variance – level 1 variance remained the same. For level 2 variance:

```
. di (exp(base[1,2])^2 - exp(e(b)[1,3])^2)/exp(base[1,2])^2
.2307748
```

So 23% of school-level variance in math achievement was explained by type of school.

Model 4. Means as outcomes model with level 1 covariate

As a next step, we can add level-1 covariates to this means-as-outcomes model. These level-1 variables can be added as fixed effects (i.e., assuming that the effects of these covariates are the same for all schools –that's what we did in model 1) or as random effects (i.e., assuming that the effects of level 1 variables vary across schools – that's what we did in model 2). We will right away opt for a more complex option, assuming that the effects of level 1 variable – SES – vary across schools.

LEVEL 1 MODEL

$$\text{MATHACH}_{ij} = \beta_{0j} + \beta_{1j}(\text{SES}_{ij}) + r_{ij}$$

LEVEL 2 MODEL

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

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

```
. mixed mathach ses sector || id: ses, cov(unstr)

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160
                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67
                                                Wald chi2(2)      =     496.29
Log likelihood = -23298.696                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     mathach | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         ses |   2.387618   .1173637    20.34   0.000     2.157589    2.617647
      sector |   2.537683   .3420706     7.42   0.000     1.867236    3.208129
       _cons |    11.4742   .2298197    49.93   0.000     11.02376    11.92464
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                    var(ses) |   .4181073   .2319119      .1409779    1.240008
                  var(_cons) |   3.895906   .5767815      2.914672    5.207475
              cov(ses,_cons) |   .7110193   .3125759      .0983819    1.323657
-----------------------------+------------------------------------------------
               var(Residual) |   36.80272   .6283463      35.59156    38.05509
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 341.24            Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

Now the intercept is the value for average SES student in a public school: 11.47. The value for an average-SES Catholic school student is 2.53 units higher: 11.47+2.54=14.01

Further, one unit increase in SES is associated with 2.39 units increase in math score (for an average school). We could check the variance component significance again to see whether SES effects vary significantly across schools now that we also control for the schools' Catholic vs public status.

```
. qui mixed mathach ses sector || id: ses, cov(unstr)

. est store sector

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df        AIC        BIC
-------------+---------------------------------------------------------------
      sector |      7,185          .   -23298.7       7   46611.39   46659.55
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.

. qui mixed mathach ses sector || id:

. lrtest sector .

Likelihood-ratio test
Assumption: . nested within sector

 LR chi2(2) =   9.04
Prob > chi2 = 0.0109

Note: The reported degrees of freedom assumes the null hypothesis is not on the
boundary of the parameter space. If this is not true, then the reported test is
conservative.

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df        AIC        BIC
-------------+---------------------------------------------------------------
           . |      7,185          .  -23303.22       5   46616.44   46650.83
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.
```

LR test indicates that SES slope variance and the covariance between intercepts and slopes are now jointly significant, even though BIC values still favor the model without random slopes.

Model 5. Intercepts and Slopes as outcomes (a.k.a. Cross-level Interactions model)

Since LR test indicated that there is some variance in SES effects across schools, we will try to explain it – we'll explore whether this variation can be attributed to the type of school – public vs Catholic (SECTOR variable).

LEVEL 1 MODEL

$$\text{MATHACH}_{ij} = \beta_{0j} + \beta_{1j}(\text{SES}_{ij}) + r_{ij}$$

LEVEL 2 MODEL

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

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

This type of model allows us to explain the variation in both intercepts and slopes. Sometimes, it's called cross-level interactions model because we make the effect of level-1 variables (SES) dependent upon the value of level-2 variables (in this case, SECTOR).

```
. mixed mathach c.ses##i.sector || id: ses, cov(unstr)

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160
                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67
                                                Wald chi2(3)      =     621.52
Log likelihood = -23281.589                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     mathach | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         ses |   2.959745   .1429495    20.70   0.000     2.679569    3.239921
    1.sector |   2.128714   .3434358     6.20   0.000     1.455593    2.801836
             |
sector#c.ses |
          1  |  -1.312909   .2153851    -6.10   0.000    -1.735056   -.8907615
             |
       _cons |   11.75256   .2301634    51.06   0.000     11.30145    12.20368
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                    var(ses) |    .073949   .0888935      .0070099    .7801067
                  var(_cons) |    3.75385   .5494602      2.817633    5.001144
             cov(ses,_cons) |   .5268683   .3298837     -.1196919    1.173429
-----------------------------+------------------------------------------------
               var(Residual) |   36.77872   .6226819      35.57831    38.01963
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 343.64                Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

In terms of fixed effects, the difference between this model and the previous one is the introduction of the effect of SECTOR on the slope of SES, which is represented as an interaction term between SECTOR and SES. That is, the effect of SES for public schools is 2.96 per one unit increase in SES; but for Catholic schools, the effect of SES is (2.96-1.31)=1.65 per one unit increase in SES. So students' math scores are more sensitive to their SES in public schools than in Catholic schools. The output shows that the effect of SES in public school (2.96) is significantly different from 0; if we wanted to know whether the effect of SES in Catholic schools is also statistically significant, we would change the omitted category of SECTOR:

```
. mixed mathach c.ses##ib1.sector || id: ses, cov(unstr)

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160
                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67
                                                Wald chi2(3)      =     621.52
Log likelihood = -23281.589                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     mathach | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         ses |   1.646837   .1611093    10.22   0.000     1.331068    1.962605
    0.sector |  -2.128714   .3434358    -6.20   0.000    -2.801836   -1.455593
             |
sector#c.ses |
          0  |   1.312909   .2153851     6.10   0.000     .8907615    1.735056
             |
       _cons |   13.88128   .2548979    54.46   0.000     13.38169    14.38087
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
id: Unstructured             |
                    var(ses) |   .0739491   .0669179      .0125506     .435714
                  var(_cons) |    3.75385    .540016      2.831561    4.976544
              cov(ses,_cons) |   .5268683   .2495518      .0377557    1.015981
-----------------------------+------------------------------------------------
               var(Residual) |   36.77872   .6212355      35.58105     38.0167
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 343.64                Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

The main effect of SES is now for Catholic schools; it's 1.65 and it is indeed statistically significant.

We can also examine the amount of variance in SES slopes explained by SECTOR: the unconditional variance in SES slopes was 0.42 (in Model 4), and the variance in this model (controlling for SECTOR x SES interaction) is only 0.07.

```
. di (0.42-0.07)/0.42
.83333333
```

Let's check again whether the remaining variation in SES slopes across schools is significant:

```
. qui mixed mathach c.ses##i.sector || id: ses, cov(unstr)

. est store interaction

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df        AIC        BIC
```

```
-------------+---------------------------------------------------------------
 interaction |      7,185          .  -23281.59       8   46579.18   46634.22
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.

. qui mixed mathach c.ses##i.sector || id:

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df        AIC        BIC
-------------+---------------------------------------------------------------
           . |      7,185          .  -23284.09       6   46580.18   46621.46
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.

. lrtest interaction .

Likelihood-ratio test
Assumption: . nested within interaction

 LR chi2(2) =   5.00
Prob > chi2 = 0.0820

Note: The reported degrees of freedom assumes the null hypothesis is not on the
boundary of the
      parameter space. If this is not true, then the reported test is conservative.
```

Both LR test and BIC indicate that SES slope variance and covariance are no longer jointly significant. Therefore, we could run this model as a model with nonrandomly varying slopes.

Model 6. Model with Nonrandomly Varying Slopes.

LEVEL 1 MODEL

$$\text{MATHACH}_{ij} = \beta_{0j} + \beta_{1j}(\text{SES}_{ij}) + r_{ij}$$

LEVEL 2 MODEL

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

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

```
. mixed mathach c.ses##i.sector || id:

Mixed-effects ML regression                     Number of obs     =      7,185
Group variable: id                              Number of groups  =        160
                                                Obs per group:
                                                              min =         14
                                                              avg =       44.9
                                                              max =         67
                                                Wald chi2(3)      =     617.69
Log likelihood = -23284.091                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     mathach | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         ses |   2.953382   .1405383    21.01   0.000     2.677932    3.228832
```

```
    1.sector |   2.137317   .3389169     6.31   0.000     1.473052    2.801582
             |
sector#c.ses |
          1  |  -1.312292   .2118813    -6.19   0.000    -1.727572   -.8970123
             |
       _cons |   11.79835   .2269266    51.99   0.000     11.35358    12.24312
------------------------------------------------------------------------------
------------------------------------------------------------------------------
  Random-effects parameters  |   Estimate   Std. err.     [95% conf. interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                  var(_cons) |   3.629644   .5210396      2.739511    4.809002
-----------------------------+------------------------------------------------
               var(Residual) |   36.83112   .6219293      35.63211    38.07047
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 338.63        Prob >= chibar2 = 0.0000
```

Note that we are able to model how sector shapes SES, but we do not allow any other variation in SES slopes because there is no significant variation beyond that accounted for by sector.