September 13, 2007
SC705: Advanced Statistics
Instructor: Natasha Sarkisian
Class notes: Two-Level HLM Models

We continue working with High School and Beyond data (included with HLM6 – see HLM folder → Examples →Chapter 2, data files: HSB1.sav and HSB2.sav).

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

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

```
Sigma_squared =     37.03440

 Tau
 INTRCPT1,B0      4.76815

Tau (as correlations)
 INTRCPT1,B0  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, B0                        0.843
 ----------------------------------------------------

The value of the likelihood function at iteration 6 = -2.332167E+004

 The outcome variable is  MATHACH
 Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                    Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          12.657481   0.187984    67.333       159    0.000
 For      SES slope, B1
    INTRCPT2, G10           2.390199   0.105719    22.609      7183    0.000
 ----------------------------------------------------------------------------

 The outcome variable is  MATHACH
 Final estimation of fixed effects
```

```
(with robust standard errors)
----------------------------------------------------------------------------
                                       Standard                    Approx.
   Fixed Effect         Coefficient   Error       T-ratio   d.f.     P-value
----------------------------------------------------------------------------
For       INTRCPT1, B0
   INTRCPT2, G00          12.657481   0.187330    67.568       159    0.000
For      SES slope, B1
   INTRCPT2, G10           2.390199   0.119309    20.034      7183    0.000
----------------------------------------------------------------------------

Final estimation of variance components:
-----------------------------------------------------------------------------
Random Effect           Standard      Variance     df    Chi-square  P-value
                        Deviation     Component
-----------------------------------------------------------------------------
INTRCPT1,       U0        2.18361       4.76815   159    1037.09077    0.000
 level-1,       R         6.08559      37.03440
-----------------------------------------------------------------------------

Statistics for current covariance components model
--------------------------------------------------
Deviance                       = 46643.331427
Number of estimated parameters = 2
```

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=4.76815/(4.76815+37.03440)= .11406362

The decrease in ρ from .18035673 to .11406362 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 Luke book, p.35-37)ЮО

(8.61431 - 4.76815)/8.61431 = .44648498

(39.14831 - 37.03440)/ 39.14831 = .05399748

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:
(39.14831+8.61431-4.76815-37.03440)/(39.14831+8.61431)= .12478524
So students' SES explained 12% of the total variance in math achievement.

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

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

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 ($\tau_{01} = \tau_{10}$). Note that covariance value indicates how much intercepts and slopes covary: in our example (below), 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).

```
Sigma_squared =     36.82835

 Tau
 INTRCPT1,B0      4.82978      -0.15399
      SES,B1     -0.15399       0.41828
```

```
Tau (as correlations)
 INTRCPT1,B0  1.000 -0.108
      SES,B1 -0.108  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, B0                        0.797
       SES, B1                        0.179
----------------------------------------------------
The value of the likelihood function at iteration 21 = -2.331928E+004

 The outcome variable is  MATHACH

 Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                       Standard                   Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          12.664935   0.189874    66.702       159    0.000
 For      SES slope, B1
    INTRCPT2, G10           2.393878   0.118278    20.240       159    0.000
 ----------------------------------------------------------------------------

 The outcome variable is  MATHACH

 Final estimation of fixed effects
 (with robust standard errors)
 ----------------------------------------------------------------------------
                                       Standard                   Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          12.664935   0.189251    66.921       159    0.000
 For      SES slope, B1
    INTRCPT2, G10           2.393878   0.117697    20.339       159    0.000
 ----------------------------------------------------------------------------

 Final estimation of variance components:
 -----------------------------------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 -----------------------------------------------------------------------------
 INTRCPT1,       U0        2.19768       4.82978   159     905.26472    0.000
      SES slope, U1        0.64675       0.41828   159     216.21178    0.002
  level-1,       R         6.06864      36.82835
 -----------------------------------------------------------------------------

 Statistics for current covariance components model
 --------------------------------------------------
 Deviance                       = 46638.560929
 Number of estimated parameters = 4
```

Here, like in the previous model, the math achievement for someone with average SES (SES=0) is 12.66; each unit increase in SES is associated with 2.39 units increase in math achievement. But, examining variance components, we notice that there is a significant variation in slopes (p-value =.002) – this means that SES effects vary across schools, so 2.39 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 36.82835, but level 2 component is (4.82978+0.41828+2*-0.15399)= 4.94008.

Note that in addition to the average reliability of school means, we now also have an estimate of reliability for the effect of SES, and it is much lower: .179.   It is normal that the reliability of slopes is much lower than that of intercepts. The precision of estimation of the intercept (which in this case is a school mean) depends only on the sample size within each school. The precision of estimation of the slope depends both on the sample size and on the variability of SES within that school. Schools that are homogeneous with respect to SES will exhibit slope estimation with poor precision. But the average reliability of the slopes is relatively low because the true slope variance across schools is much smaller than the variance of the true means.

Note that low reliabilities do not invalidate the HLM analysis, but very low reliabilities (typically < .10) often indicate that a random coefficient might be considered fixed (i.e., the same across groups) in subsequent analyses.

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!

```
Sigma_squared =     39.15135

 Tau
 INTRCPT1,B0      6.67771

Tau (as correlations)
 INTRCPT1,B0  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, B0                        0.877
 ----------------------------------------------------
The value of the likelihood function at iteration 4 = -2.353915E+004

 The outcome variable is  MATHACH
 Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                       Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          11.393043   0.292887    38.899       158    0.000
      SECTOR, G01           2.804889   0.439142     6.387       158    0.000
```

```
-----------------------------------------------------------------------------

 The outcome variable is  MATHACH
 Final estimation of fixed effects
 (with robust standard errors)
 ----------------------------------------------------------------------------
                                       Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          11.393043   0.292258    38.983       158    0.000
      SECTOR, G01           2.804889   0.435823     6.436       158    0.000
 ----------------------------------------------------------------------------

 Final estimation of variance components:
 -----------------------------------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 -----------------------------------------------------------------------------
 INTRCPT1,       U0        2.58413       6.67771   158    1296.76559    0.000
  level-1,       R         6.25710      39.15135
 -----------------------------------------------------------------------------

 Statistics for current covariance components model
 --------------------------------------------------
 Deviance                       = 47078.295826
 Number of estimated parameters = 2
```

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:

(8.61431 - 6.67771)/8.61431 = .22481197

So 22% 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

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

LEVEL 2 MODEL

$\beta_{0j} = \gamma_{00} + \gamma_{01}(SECTOR_j) + u_{0j}$

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

```
Sigma_squared =     36.79508

 Tau
 INTRCPT1,B0      3.96459       0.71641
      SES,B1      0.71641       0.44990

Tau (as correlations)
 INTRCPT1,B0  1.000  0.536
      SES,B1  0.536  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, B0                        0.765
       SES, B1                        0.189
 ----------------------------------------------------

The value of the likelihood function at iteration 21 = -2.330093E+004
 The outcome variable is  MATHACH

 Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                        Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          11.476646   0.231587    49.557       158    0.000
      SECTOR, G01           2.533835   0.344798     7.349       158    0.000
 For      SES slope, B1
    INTRCPT2, G10           2.385451   0.118329    20.160       159    0.000
 ----------------------------------------------------------------------------

 The outcome variable is  MATHACH

 Final estimation of fixed effects
 (with robust standard errors)
 ----------------------------------------------------------------------------
                                        Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          11.476646   0.225026    51.001       158    0.000
      SECTOR, G01           2.533835   0.352411     7.190       158    0.000
 For      SES slope, B1
    INTRCPT2, G10           2.385451   0.119008    20.044       159    0.000
 ----------------------------------------------------------------------------

 Final estimation of variance components:
 -----------------------------------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 -----------------------------------------------------------------------------
 INTRCPT1,       U0        1.99113       3.96459   158     766.83844    0.000
```

```
     SES slope, U1        0.67075       0.44990   159     216.12223    0.002
 level-1,       R         6.06589      36.79508
-----------------------------------------------------------------------------

Statistics for current covariance components model
--------------------------------------------------
Deviance                       = 46601.861400
Number of estimated parameters = 4
```

Now the intercept is the value for average SES student in a public school: 11.48. The value for an average-SES Catholic school student is 2.53 units higher: 11.45+2.53=13.98
Further, one unit increase in SES is associated with 2.39 units increase in math score. But there is still significant variation across schools in intercepts, and there is also significant variation in SES slopes – so SES doesn't have the same effect across schools.

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

Next, we will try to explain this variation in SES effects across schools – we'll explore whether this variation can be attributed to the type of school – public vs Catholic.

LEVEL 1 MODEL

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

LEVEL 2 MODEL

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(SECTOR_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(SECTOR_j) + u_{1j}$$

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).

```
Sigma_squared =     36.76311

 Tau
 INTRCPT1,B0      3.83295       0.54112
      SES,B1      0.54112       0.12988

Tau (as correlations)
 INTRCPT1,B0  1.000  0.767
      SES,B1  0.767  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, B0                        0.759
       SES, B1                        0.064
 ----------------------------------------------------
The value of the likelihood function at iteration 198 = -2.328373E+004

 The outcome variable is  MATHACH
 Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                        Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
```

```
   INTRCPT2, G00          11.750237   0.232241    50.595       158    0.000
     SECTOR, G01           2.128611   0.346651     6.141       158    0.000
For      SES slope, B1
   INTRCPT2, G10           2.958798   0.145460    20.341       158    0.000
     SECTOR, G11          -1.313096   0.219062    -5.994       158    0.000
----------------------------------------------------------------------------

The outcome variable is  MATHACH
Final estimation of fixed effects
(with robust standard errors)
----------------------------------------------------------------------------
                                      Standard                 Approx.
   Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
----------------------------------------------------------------------------
For       INTRCPT1, B0
   INTRCPT2, G00          11.750237   0.218675    53.734       158    0.000
     SECTOR, G01           2.128611   0.355697     5.984       158    0.000
For      SES slope, B1
   INTRCPT2, G10           2.958798   0.144092    20.534       158    0.000
     SECTOR, G11          -1.313096   0.214271    -6.128       158    0.000
----------------------------------------------------------------------------
Final estimation of variance components:
-----------------------------------------------------------------------------
Random Effect           Standard      Variance     df    Chi-square  P-value
                        Deviation     Component
-----------------------------------------------------------------------------
INTRCPT1,       U0        1.95779       3.83295   158     756.04082    0.000
     SES slope, U1        0.36039       0.12988   158     178.09113    0.131
 level-1,       R         6.06326      36.76311
-----------------------------------------------------------------------------

Statistics for current covariance components model
--------------------------------------------------
Deviance                       = 46567.464841
Number of estimated parameters = 4
```

In terms of fixed effects, the difference between this model and the previous one is the introduction of the effect of SECTOR on SES, which can be interpreted 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.

We can also examine the amount of variance in SES slopes explained by SECTOR: the unconditional variance in SES slopes was 0.44990, and the variance in this model (controlling for SECTOR) is only 0.12988.

(0.44990-0.12988)/0.44990 = .71131363

So SECTOR explained 71% of between-school variance in effects of SES on math achievement. Also note that now that we controlled for sector, the variation in SES slopes across schools is no longer significant. Therefore, we could run this model as a model with nonrandomly varying slopes.

Model 6. Model with Nonrandomly Varying Slopes.

LEVEL 1 MODEL

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

LEVEL 2 MODEL

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(SECTOR_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(SECTOR_j)$$

```
 Sigma_squared =     36.84019

 Tau
 INTRCPT1,B0      3.69423

Tau (as correlations)
 INTRCPT1,B0  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, B0                        0.808
 ----------------------------------------------------
The value of the likelihood function at iteration 6 = -2.328616E+004

 The outcome variable is  MATHACH
 Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                       Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          11.797994   0.228514    51.629       158    0.000
      SECTOR, G01           2.138170   0.341344     6.264       158    0.000
 For      SES slope, B1
    INTRCPT2, G10           2.951177   0.140609    20.989      7181    0.000
      SECTOR, G11          -1.312849   0.211996    -6.193      7181    0.000
 ----------------------------------------------------------------------------

 The outcome variable is  MATHACH
 Final estimation of fixed effects
 (with robust standard errors)
 ----------------------------------------------------------------------------
                                       Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          11.797994   0.214976    54.880       158    0.000
      SECTOR, G01           2.138170   0.350028     6.109       158    0.000
 For      SES slope, B1
    INTRCPT2, G10           2.951177   0.143779    20.526      7181    0.000
      SECTOR, G11          -1.312849   0.212824    -6.169      7181    0.000
 ----------------------------------------------------------------------------

 Final estimation of variance components:
 -----------------------------------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 -----------------------------------------------------------------------------
```

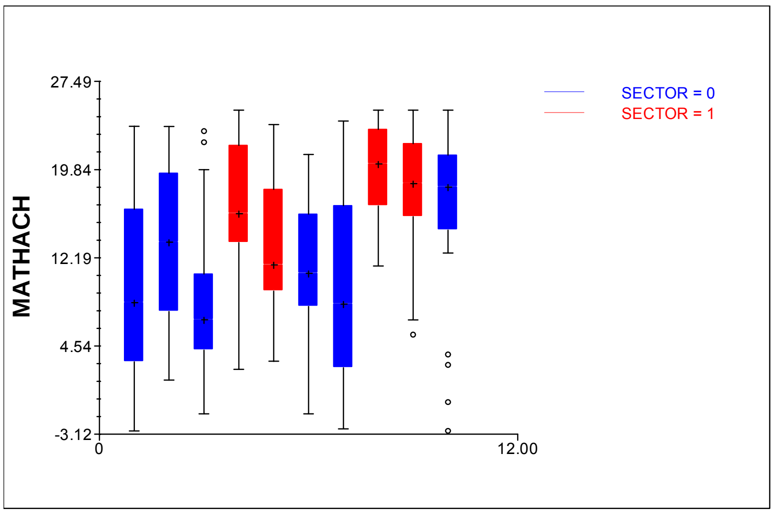
```
INTRCPT1,       U0        1.92204       3.69423   158      837.19099     0.000
 level-1,       R         6.06961      36.84019
-----------------------------------------------------------------------------

Statistics for current covariance components model
--------------------------------------------------
Deviance                        = 46572.326387
Number of estimated parameters = 2
```

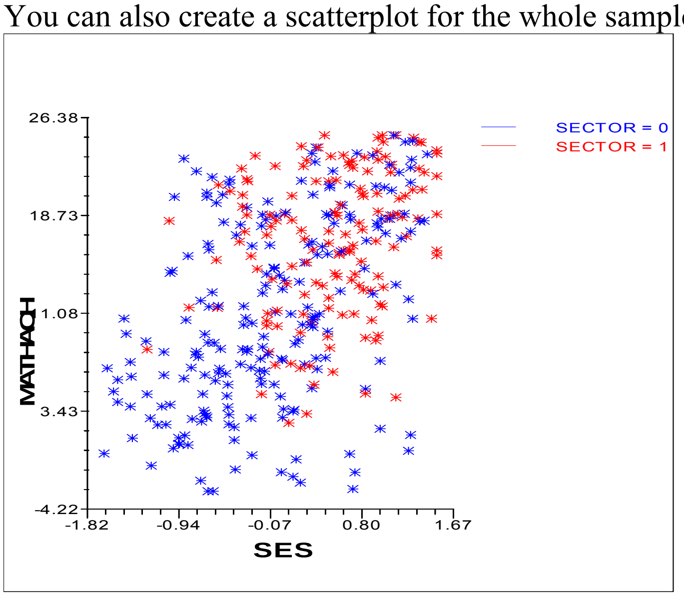
Note that we are able to model how sector shapes SES, but we do not allow any other variation in SES slopes because there is no significant variation beyond that accounted for by sector.

Using graphs to examine the data

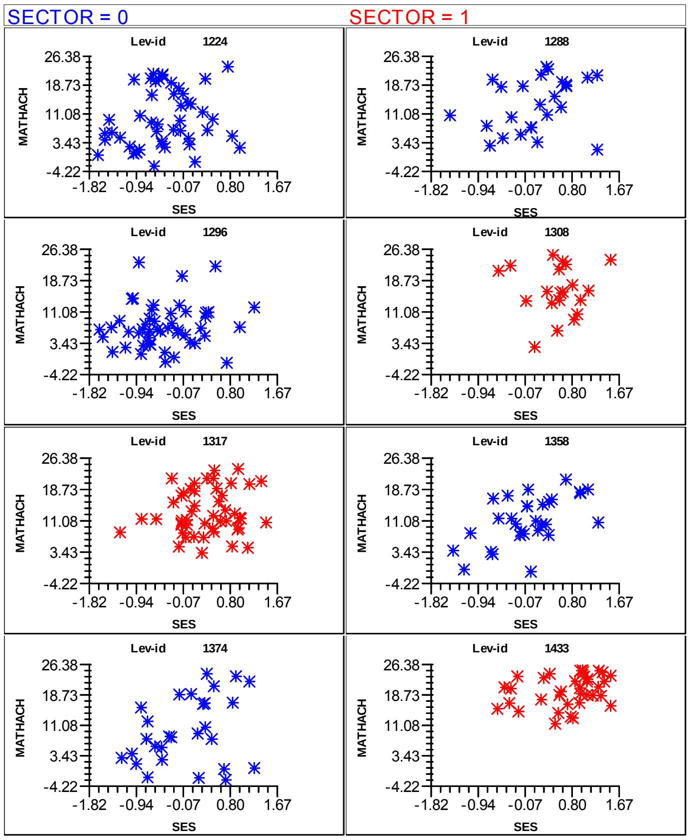
HLM has some limited graphing capabilities allowing you to examine the data before starting to build models. You can examine your data by creating boxplots for a variable, e.g., your dependent variable, by group. You can also mark these groups according to one level-2 variable:



You can also create a scatterplot for the whole sample by values of a level-2 variable:



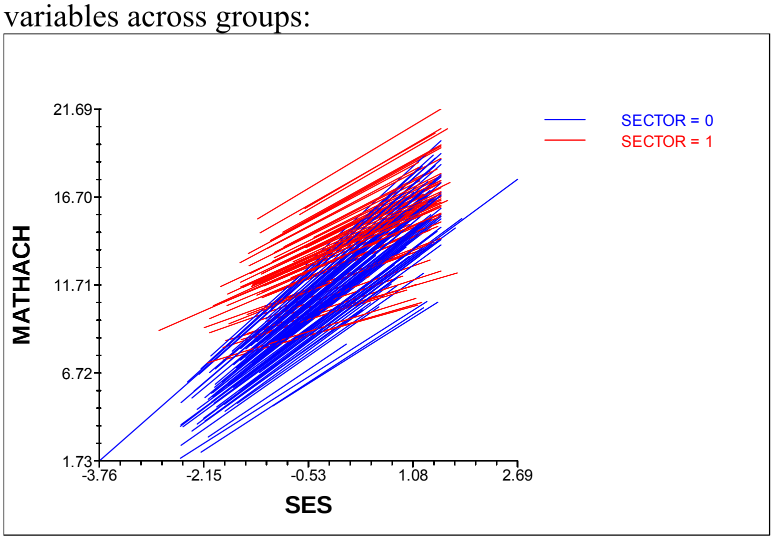
Or you can create scatterplots separating groups and colorcoding them by a level-2 variable:



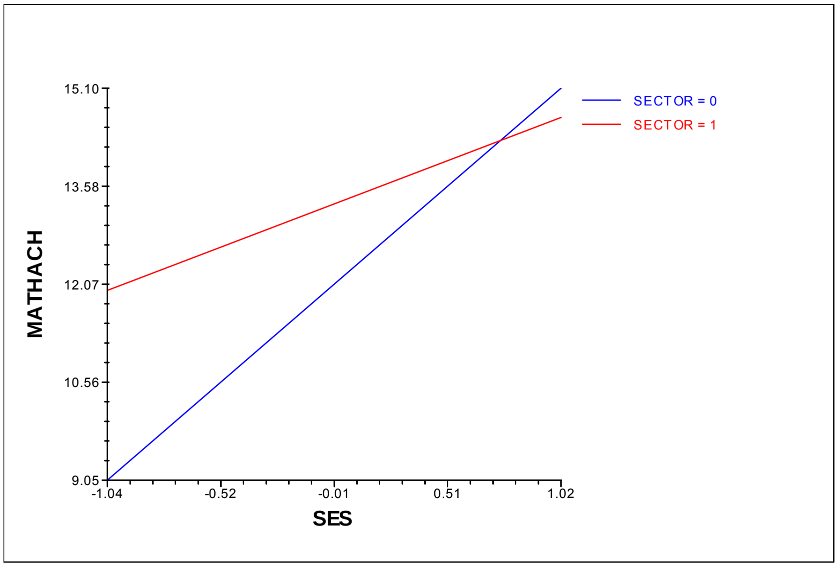
Graphing Equations

Graphs can also be used to better illustrate and understand the models you estimate. These can greatly assist in interpreting the findings. HLM offers a range of such graphing options.

E.g., Graph Equations → Level 1 equation graphing. Here you can examine slopes for level-1 variables across groups:



Or you can graph the relationships based on the fixed effects in your last model using Graph Equations → Model graphs.   Here, you have a range of options. For example, you can look at how level 1 slopes vary depending on values of level 2 variables (if you have a cross-level interaction in your model). You need to select a level-1 variable as you X, and level-2 variable as Z focus:



Or you can examine how predicted values vary by level of both level-1and level-2 variables by selecting level-2 variable as your X, and level-1 as your Z-focus: