Dataset 1 - Strength of HLM over OLS.
For use with HLM 6

In this dataset, users are shown that sometimes the result of an HLM analysis can be exactly the opposite from the results of Ordinary Least Squares (OLS). HLM 6 uses datasets that are disaggregated at the appropriate level. The two datasets needed for this analysis are "eg2_L1.sav" for the level-1 data and "eg2_L2.sav" for the level-2 data. Both datasets are in SPSS format.

In this example, the dependent variable is a science achievement test score measured at the individual level. The independent variable called "urban" is a composite score from scores on student SES and other markers of student "urbanicity". Thus, high student scores on "urban", represent a student who has low SES and also has many of the markers associated with students in a highly urban setting. The data represent 160 students nested in 16 schools (10 students in each school). The level 2 grouping structure is identified by the variable "group".

First, let's look at the results from the OLS procedure. These are computed in S-PLUS.

```
Call: lm(formula = SCIENCE ~ URBAN, data = example.data)
Residuals:
    Min     1Q Median    3Q   Max
 -5.336 -2.129 0.4919 2.043 5.009

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)  -1.2511   0.5937    -2.1072   0.0367
      URBAN   0.8276   0.0386    21.4248   0.0000

Residual standard error: 2.592 on 158 degrees of freedom
Multiple R-Squared: 0.7439
F-statistic: 459 on 1 and 158 degrees of freedom, the p-value is 0

Correlation of Coefficients:
      (Intercept)
URBAN -0.9386
```

If we were to trust these results without looking at further HLM estimates, we might assume that as students become more "urban", their scores on the science achievement variable tend to increase (c.f., slope coefficient of 0.8276). As we will see from the HLM analysis, this would be an egregious error.

Roberts, J. K. (2005, April). Datasets illustrating specific strengths of hierarchical linear modeling. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

We first fit the null model, also called the unconditional model, in which just the intercept is estimated in a random coefficients model. The full model can be represented as:

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

with a level-1 model of:

$$y_{ij} = \beta_{0j} + e_{ij}$$

where y_{ij} represents the scores for student "*i*" in school "*j*" on the dependent variable, β_{0j} is the mean science achievement for school "*j*", and e_{ij} are the student-level random deviates around school "*j*'s" mean. Further extrapolated, we can see that for each student at level-1, their score is represented by the following equations:

$$\begin{cases} y_{11} = \beta_1 + e_{11} \\ y_{21} = \beta_1 + e_{21} \\ \quad\cdots \\ y_{ij} = \beta_j + e_{ij} \end{cases} .$$

The level-2 model would then be:

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

where β_{0j} is the mean science achievement for school "*j*", γ_{00} is the overall grand intercept, and u_{0j} is school "*j*'s" random deviate around the grand mean.

To create the .MDM file (be sure to click on ***Stat Package Input***) if using SPSS data (.SAV) files as input.

Figure 1

Creating a Multivariate Data Matrix (MDM) File

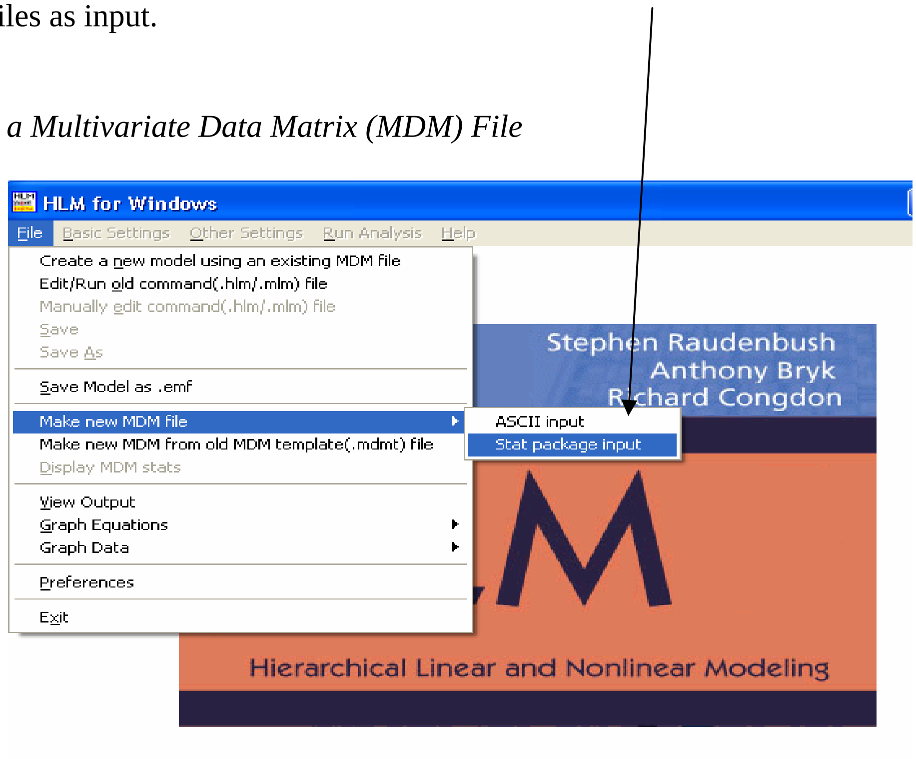
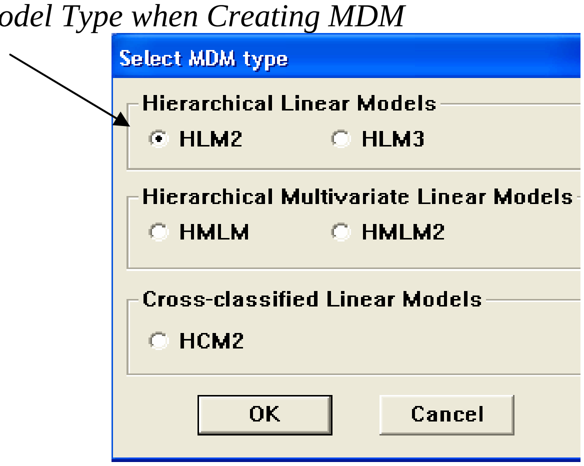


Figure 2

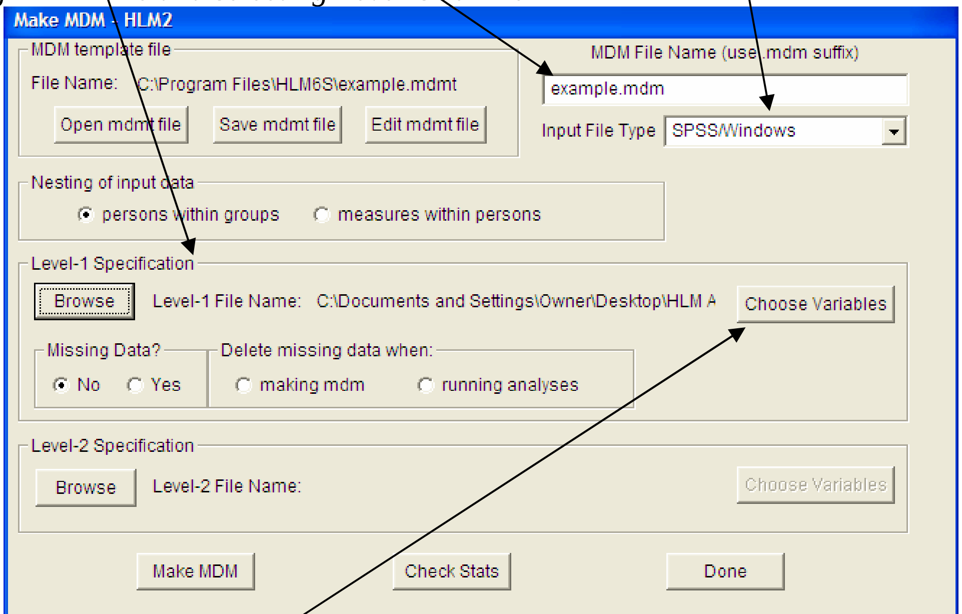
Selecting HLM Model Type when Creating MDM



First name the MDM file for later use, check the correct data file type and select the Level One file

Figure 3

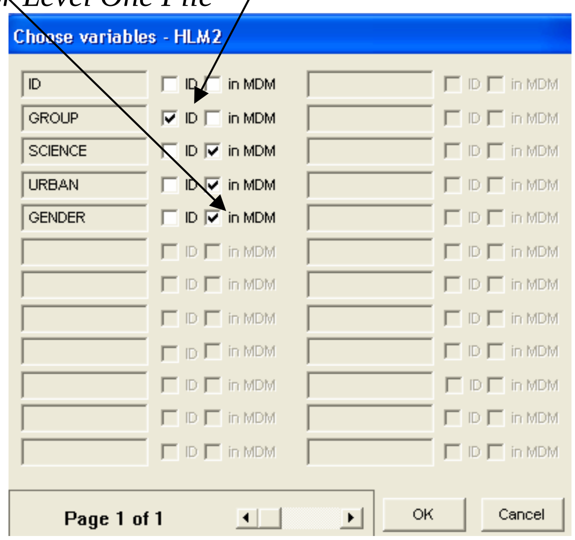
Naming MDM File and selecting Level One File



Then click on ***Choose Variables*** to identify the ID variable, dependent variable and level one predictors

Figure 4

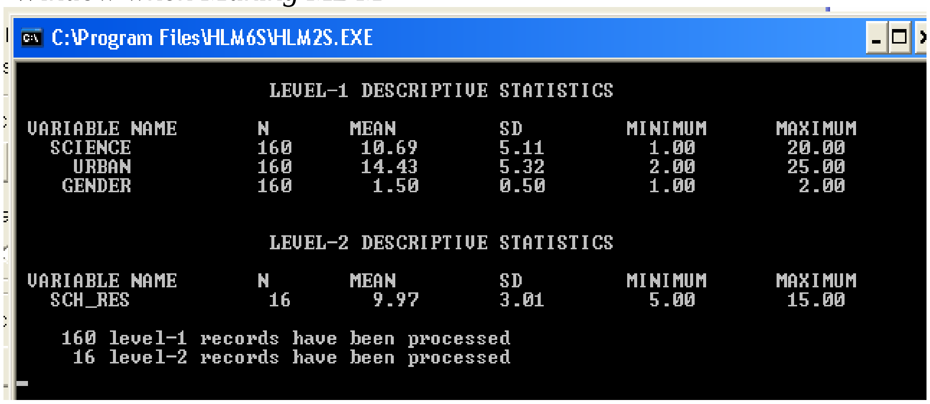
Choosing Variables for Level One File



Identify missing data as necessary (back on the ***Make MDM*** window). Next select the Level two data file and then click on ***Choose Variables*** for the Level two file. You still need to identify the linking level two id and any Level two predictors.

Next click on ***Save MDMT*** file, giving a name for the file. Then click on ***Make MDM*** and you will see an MS-DOS window that (only briefly) appears (see Figure 5). This window simply lists some descriptive statistics for the variables of interest:

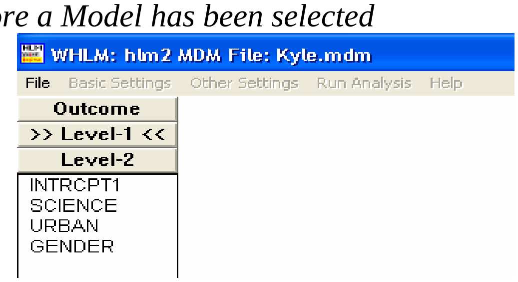
Figure 5
MS-DOS Window when Making MDM



These same statistics will appear again once you click on ***Check Stats***. Lastly, click on ***Done***.

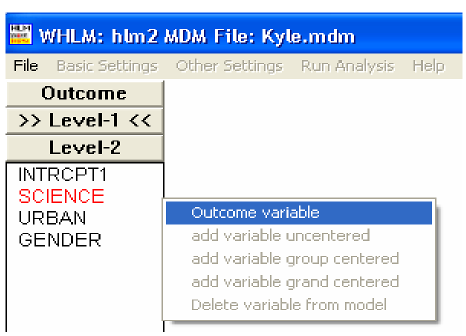
Then you will see a blank HLM screen:

Figure 6
Blank HLM Screen Before a Model has been selected



First select the outcome variable, ***Science***:

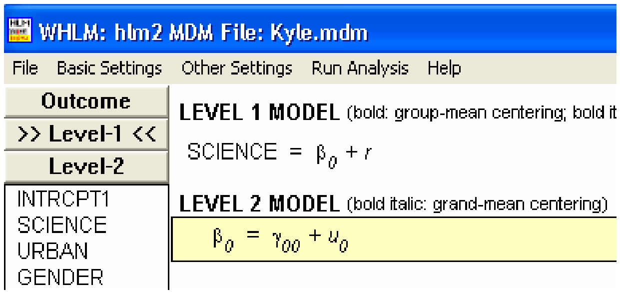
Figure 7
Identifying the Outcome



Which results in Figure 8:

Figure 8

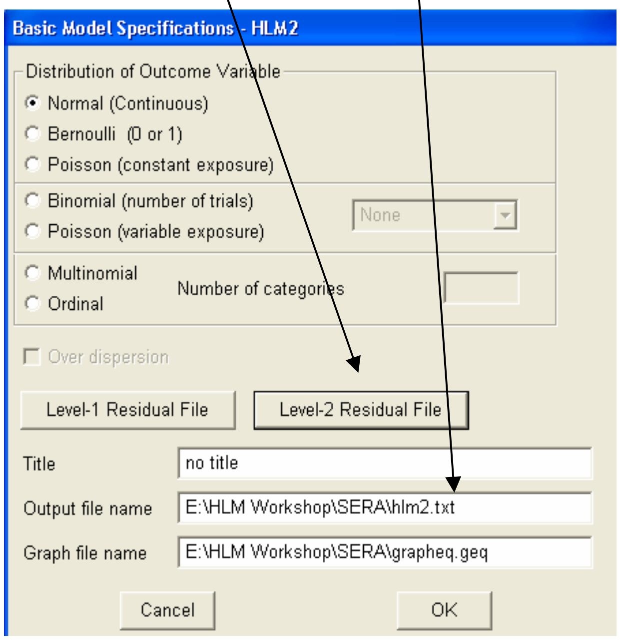
The Two-Level Unconditional Model in HLM



Now click on ***Basic Settings*** to provide a specific filename for the model you are estimating (change for each model so you can have output file for each model's estimates!). In ***Basic Settings***, you can also request Level 2 Residuals.

Figure 9

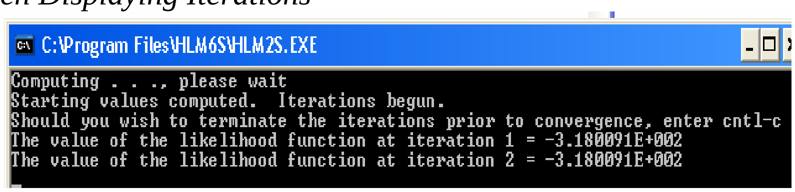
Basic Settings Screen



Then click on ***File*** ⇒ ***Save As*** to save the model file which you must do before you can click on ***Run Analysis*** to estimate your model's parameters. Again you will see an MS-DOS screen. This one displays the iterations required till convergence.

Figure 10

MS-DOS Screen Displaying Iterations



You can then view the output by clicking on ***File*** ⇒ ***View Output***

Figure 11
Two-Level Unconditional Model Output

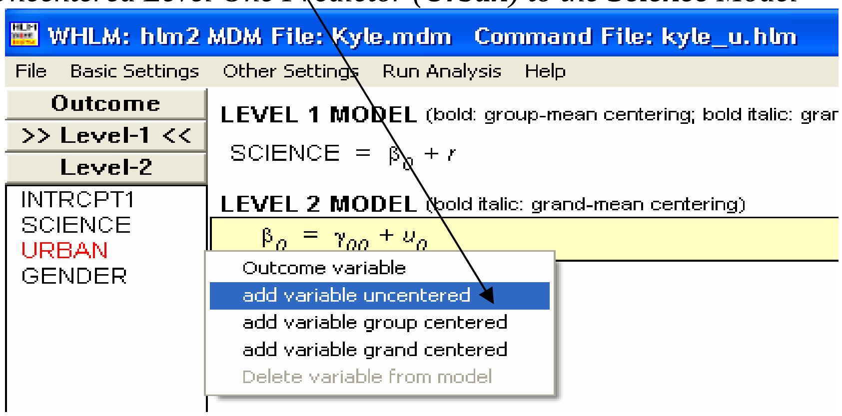
```
Program:                       HLM 6 Hierarchical Linear and Nonlinear Modeling
Authors:                       Stephen Raudenbush, Tony Bryk, & Richard Congdon
Publisher:                     Scientific Software International, Inc. (c) 2000
                                                      techsupport@ssicentral.com
                                                              www.ssicentral.com
--------------------------------------------------------------------------------
Module:      HLM2S.EXE (6.00.24282.2)
Date:        7 February 2005, Monday
Time:        22:10:19
--------------------------------------------------------------------------------


SPECIFICATIONS FOR THIS HLM2 RUN


 Problem Title: no title

 The data source for this run  = example.mdm
 The command file for this run = E:\HLM Workshop\SERA\example.hlm
 Output file name              = E:\HLM Workshop\SERA\example.txt
 The maximum number of level-1 units = 160
 The maximum number of level-2 units = 16
 The maximum number of iterations = 100
 Method of estimation: restricted maximum likelihood

Weighting Specification
-----------------------
                        Weight
                        Variable
           Weighting?   Name        Normalized?
Level 1        no
Level 2        no
Precision      no

 The outcome variable is  SCIENCE

 The model specified for the fixed effects was:
----------------------------------------------------

  Level-1                  Level-2
  Coefficients             Predictors
----------------------   ---------------
        INTRCPT1, B0      INTRCPT2, G00


The model specified for the covariance components was:
---------------------------------------------------------

        Sigma squared (constant across level-2 units)

        Tau dimensions
              INTRCPT1


Summary of the model specified (in equation format)
---------------------------------------------------

Level-1 Model

       Y = B0 + R

Level-2 Model
       B0 = G00 + U0
```

```
Iterations stopped due to small change in likelihood function

******* ITERATION 2 *******

 Sigma_squared =      1.97917

 Tau
 INTRCPT1,B0     25.53125


Tau (as correlations)
 INTRCPT1,B0  1.000

 ----------------------------------------------------
  Random level-1 coefficient   Reliability estimate
 ----------------------------------------------------
  INTRCPT1, B0                        0.992
 ----------------------------------------------------

The value of the likelihood function at iteration 2 = -3.180091E+002

 The outcome variable is  SCIENCE

 Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                       Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          10.687500   1.268098     8.428        15    0.000
 ----------------------------------------------------------------------------

 The outcome variable is  SCIENCE

 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          10.687500   1.227831     8.704        15    0.000
 ----------------------------------------------------------------------------


The robust standard errors are appropriate for datasets having a moderate to
large number of level 2 units.  These data do not meet this criterion.

 Final estimation of variance components:
 -----------------------------------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 -----------------------------------------------------------------------------
 INTRCPT1,       U0        5.05285      25.53125    15    1950.00000    0.000
  level-1,       R         1.40683       1.97917
 -----------------------------------------------------------------------------


 Statistics for current covariance components model
 --------------------------------------------------
 Deviance                       = 636.018232
 Number of estimated parameters = 2
```

To add ***Urban*** to the model, just click on ***Urban*** as in Figure 12. You will see a list of choices for options for centering of a Level One variable.

Figure 12

*Adding an Uncentered Level One Predictor (**Urban**) to the **Science** Model*



"Uncentered" means that the predictor is not transformed. "Group centered" means that scores are transformed so they represent deviations from their group means. "Grand centered" means that scores are transformed so they represent deviations from the grand mean (across the whole sample).

If ***Urban*** is chosen as an Uncentered variable, you will notice that it is modeled as a ***<u>fixed</u>*** effect:

Figure 13

***Science** Model with **Urban** as a Fixed Effect*

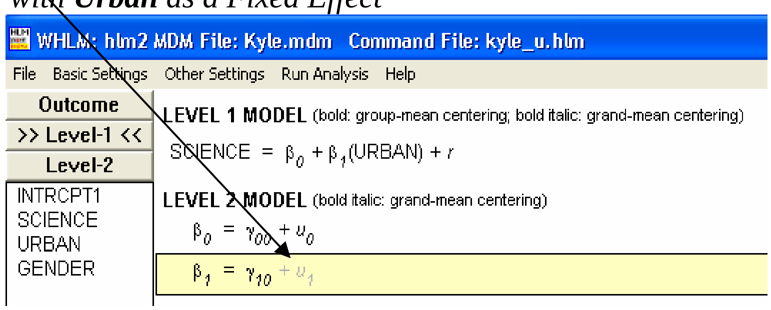


Figure 14

*Abbreviated Output for **Science** Model with Fixed **Urban***

```
Summary of the model specified (in equation format)
 ---------------------------------------------------

Level-1 Model

        Y = B0 + B1*(URBAN) + R

Level-2 Model
        B0 = G00 + U0
        B1 = G10
…
```

Final estimation of fixed effects:

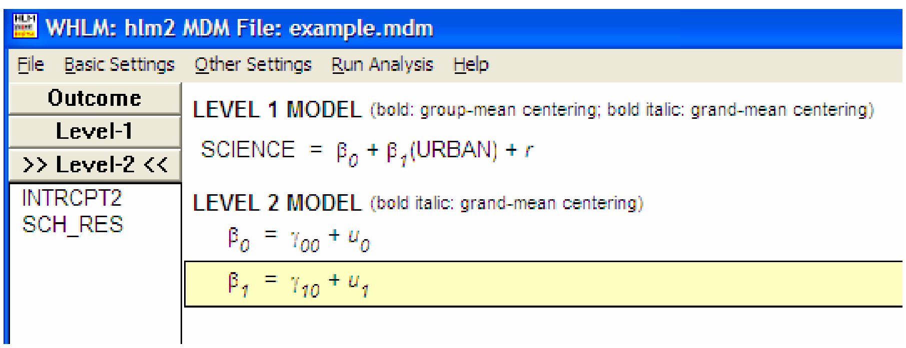
Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	22.302911	2.426310	9.192	15	0.000
For URBAN slope, B1					
INTRCPT2, G10	-0.805228	0.047999	-16.776	158	0.000

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	9.29817	86.45595	15	19792.95548	0.000
level-1,	R	0.80945	0.65521			

<u>Figure 1</u>

*Adding a random effect for **Urban***

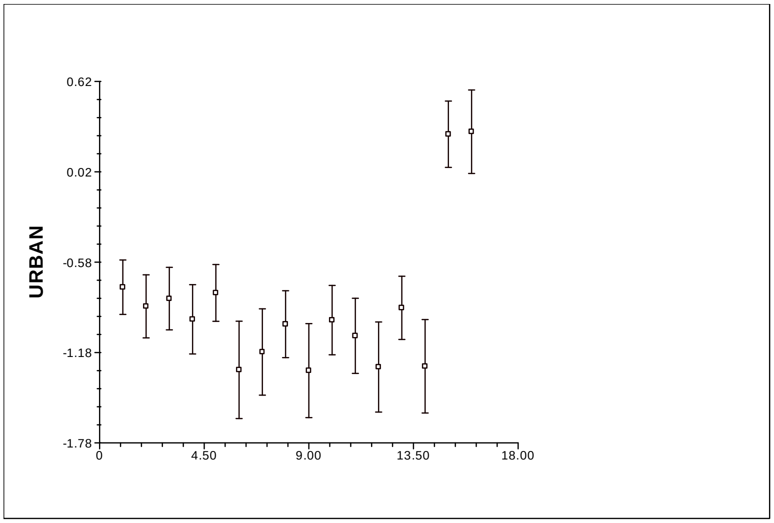


To add ***Urban*** as a Random effect (across groups), just click on the relevant Level 2 equation (the second one) and the random effect will be inserted. Click again and ***Urban*** will be a fixed effect.

Figure 16
*Abbreviated Output from Adding **Urban** as Random Effect*

```
Summary of the model specified (in equation format)
 ---------------------------------------------------

Level-1 Model

       Y = B0 + B1*(URBAN) + R

Level-2 Model
       B0 = G00 + U0
       B1 = G10 + U1…

Final estimation of fixed effects:
 ----------------------------------------------------------------------------
                                       Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00          22.391241   2.716968     8.241        15    0.000
 For    URBAN slope, B1
    INTRCPT2, G10          -0.867005   0.129808    -6.679        15    0.000
 ----------------------------------------------------------------------------


Final estimation of variance components:
 -----------------------------------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 -----------------------------------------------------------------------------
 INTRCPT1,       U0       10.65844     113.60241    15    1626.56233    0.000
    URBAN slope, U1        0.50199       0.25200    15     219.72077    0.000
  level-1,       R         0.52025       0.27066
 -----------------------------------------------------------------------------
```

Finally, we can see from the last model that the relationship between “science” and “urban” in this dataset is actually negative, meaning that as a student has higher classification of the urban composite variable, scores on the science achievement variable tend to decrease. Remember that this is exactly the opposite from the conclusion that we drew from the OLS analysis. Also, we can see from either the fitted estimates or from the above graph that schools “15” and “16” have slopes that are dramatically different from the other schools. To view the graphical display of the Empirical Bayes estimates of the slope coefficients, click on “File” → “Graph Equations” → “Level-2 EB/OLS coefficient confidence intervals”. Then select “urban” under “Level-1” and change the number of groups to be “All groups (n=16)”. Doing this will produce:



You may also observe the second-level residuals in a file called “resfil2.sav”. Simply open this file in SPSS and you can see that the Empirical Bayes estimated intercepts and slopes in the variables “ecintrcp” and “ecurban” respectively.

	ecintrcp	ecurban
1	7.038	-.747
2	8.901	-.874
3	11.669	-.823
4	15.130	-.961
5	16.186	-.785
6	26.029	-1.297
7	24.551	-1.178
8	24.894	-.993
9	31.570	-1.302
10	26.967	-.966
11	30.983	-1.071
12	36.360	-1.278
13	31.268	-.884
14	41.201	-1.273
15	12.724	.271
16	12.788	.288