HLM — An Introduction

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Multilevel Regression Modeling, 2009

HLM — An Introduction

Introduction

Today we look back at some of the analyses we did in the last lecture, and recast them in the analytic framework of the popular statistics program HLM.

The HLM Program

HLM is a popular software program that makes construction of basic multilevel models relatively straightforward. In particular, it does not require combination of models from two or more levels into a single regression model. Consequently, many find it very convenient and (relatively) easy to use, which has contributed to its popularity. In this introduction, we will revisit the models that we examined, and set them up and analyze them in HLM.

We assume that you have the HLM6 program (full or student version) installed on your computer.

Two-Level Models in HLM

HLM uses a consistent notation for its models. Since this notation is displayed while models are being specified, it is easier to see precisely what has been specified.

Note that unlike the Gelman and Hill notation, the HLM notation implicitly assumes that data are broken into files by level, and therefore finds it convenient to specify the level 2 unit explicitly in the notation.

The General Level-1 Model

Consider, for example, our radon data, in which houses are nested within counties, and at level-1 we wish to predict radon level from floor.

In this notational scheme, Y_{ij} stands for the outcome score (radon level) of the ith level-1 unit (i.e., the ith house) within the jth level-2 unit (county). So, for example, $Y_{1,13}$ would refer to the first house in the 13th county.

The General Level-1 Model

The basic model is

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + \ldots + \beta_{Qj} X_{Qij} + r_{ij} \tag{1}$$

In this model, the β_{qj} are level-1 coefficients, X_{qij} is the qth level-1 predictor for level-1 unit i within level-2 unit j. (The HLM manual refers to this as the predictor for the "ith case in unit j.") r_{ij} is the level-1 random effect, and σ^2 is the variance of r_{ij}. It is assumed that $r_{ij} \sim \mathrm{N}(0, \sigma^2)$. By giving the β's a second subscript we allow them to vary across level-2 units, so we can have variable slopes, variable intercepts, both, or neither.

The General Level-2 Model

The general level-2 model is

$$\beta_{qj} = \gamma_{q0} + \gamma_{q1} W_{1j} + \gamma_{q2} W_{2j} + \ldots + \gamma_{qS_q} W_{S_q j} + u_{qj} \qquad (2)$$

The γ's are *level-2 coefficients*, the W's are level-2 predictors, and the u's are level-2 random effects. The u's have a covariance matrix $\boldsymbol{T}$ with typical element $\tau_{qq'}$.

Special Cases

The general two-level model allows for numerous special cases. For example, we can have

1. A fixed level-1 coefficient, i.e., $\beta_{qj} = \gamma_{q0}$
2. A non-randomly varying level-1 coefficient, i.e., $\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj}$. This corresponds to the full level-2 specification without the random component.
3. A randomly varying level-1 coefficient with no level 2 predictors, i.e., $\beta_{qj} = \gamma_{q0} + u_{qj}$, or
4. The full level-2 system $\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} + u_{qj}$.

One-Way ANOVA with Random Effects – The Model

As we saw in the previous lecture, an extremely simple multilevel model has no predictors at either level-1 or level-2. In the HLM scheme, this may be written simply as follows. The level-1 model is

$$\text{RADON}_{ij} = \beta_{0j} + r_{ij} \tag{3}$$

The level-2 model is

$$\beta_{0j} = \gamma_{00} + u_{0j} \tag{4}$$

These can be combined into a single model,

$$\text{RADON}_{ij} = \gamma_{00} + u_{0j} + r_{ij} \tag{5}$$

which you can see is of the classic random-effects ANOVA form

$$y_{ij} = \mu + a_j + \epsilon_{ij} \tag{6}$$

Data Preparation and Input

HLM has limited (and somewhat disguised) data input capabilities. In practice, you will probably input most of your data as either SPSS *.sav* files, or comma-delimited ASCII files with a header containing column names. Since R writes ascii files routinely using the `write.table()` function (and the `sep = ','` option), and also has extensive data manipulation capabilities, you may find it convenient to use R to construct your HLM files.

Data Preparation and Input

The link between the level-1 and level-2 models in the HLM parameterization is the subscript j, which refers to the `county` variable. To set up the data for HLM, we need two files, one for the level-1 variables, one for the level-2 variables. Each file has to be sorted in ascending order of the ID variable.

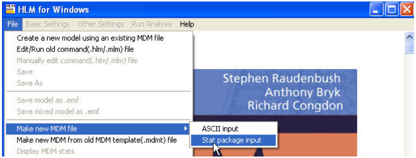
We include `county`, a log-transformed `radon`, and the `floor` predictor in the level-1 file, and `county` and `uranium` in the level-2 file.

Since (unlike R), HLM does not have built-in data transformation capabilities, we log-transform `radon` prior to saving the level-1 file.

Constructing the MDM File

Setting up the Multivariate Data Matrix (MDM) file is a key first step to using HLM2 to analyze a 2-level model problem.

Make sure you have downloaded the files **`radon1.txt`** and **`radon2.txt`** from the course website. Begin by starting up HLM. Then click on the *Make New MDM File -> Stat Package Input* menu option. (This is counterintuitive and very poor human factors design, since we are loading an ASCII file. Of course, this should be available under the ACII file input node.)

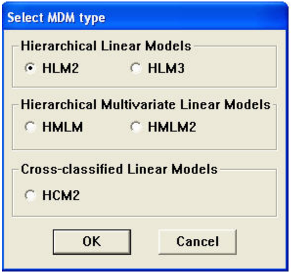


Constructing the MDM File

Next, you will be asked to select a program.

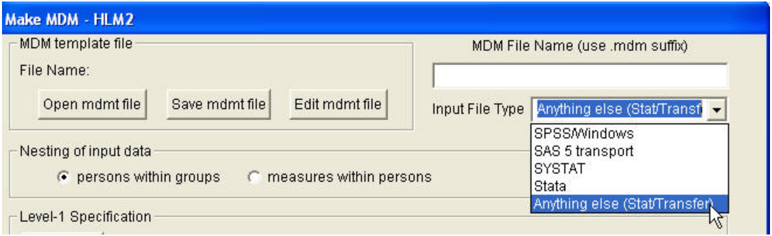
Select HLM2.

Then click on the OK button.



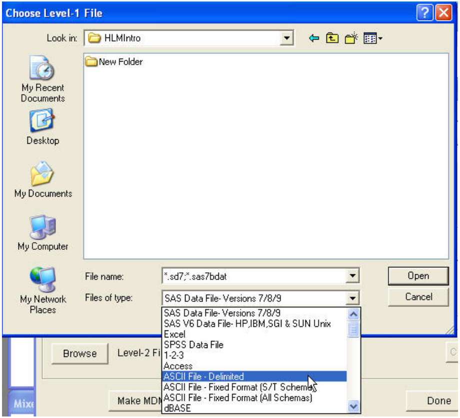
Constructing the MDM File

A large dialog box will open. Go to the drop-down list for file type, and select *Anything else (Stat/Transfer)*.



Constructing the MDM File

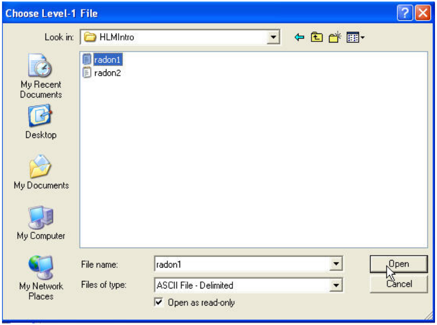
Use the drop-down box to select the delimited ASCII file type.



Constructing the MDM File

If you are not already there, go to the directory where the data files are.

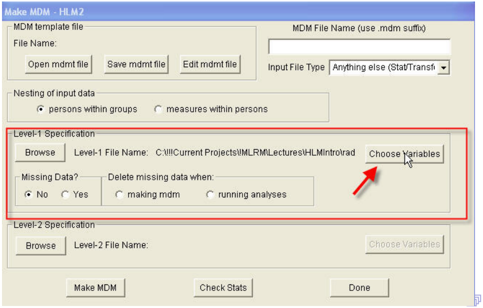
Select *radon1.txt* and click on the *Open* button.



Constructing the MDM File

Look in the dialog box for the grouping that is titled *Level-1 Specification*. I've highlighted the group in red in the picture below.

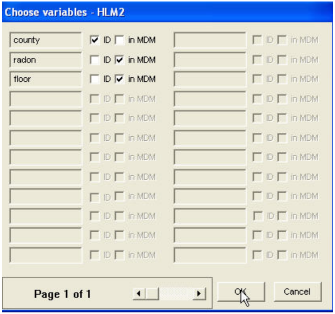
On the right side of that grouping is a button *Choose Variables*.



Click on it.

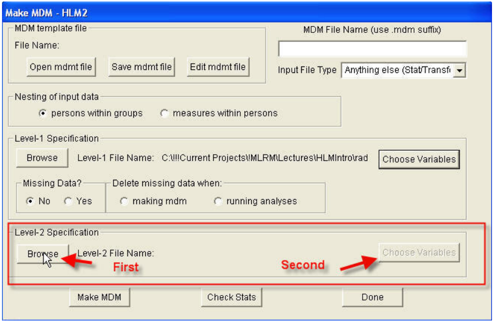
Constructing the MDM File

A dialog box will open up that will allow you to select and classify level-1 variables. The variable that spans the two levels of your model is `county`, and this variable is designated an ID variable. The variables `radon` and `floor` are in the level-1 model, so they are checked off as being in the MDM. When you are ready to exit the dialog, it should look like this:



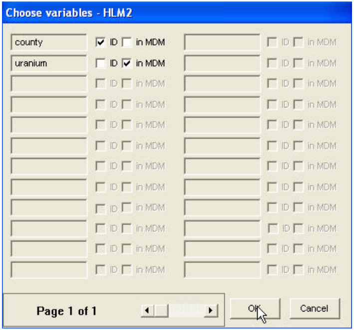
Constructing the MDM File

Next, go to the *Level-2 Specification* group and first, click on the *Browse* button. Select the file *radon2.txt*. Second, click on the *Choose Variables* button.



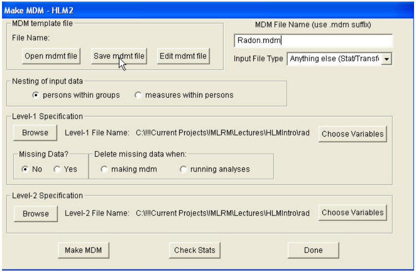
Constructing the MDM File

This will take you to another variable selection dialog. Again `county` is the ID variable, and `uranium`, a level-2 predictor, is added to the MDM file. We will not use all the variables in the MDM file in our first model, but we can re-use this file for other more complicated models.



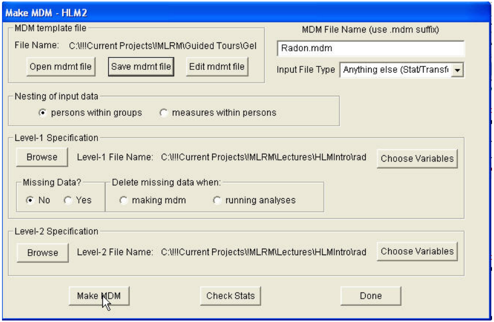
Constructing the MDM File

At this point, the HLM program again exhibits poor human factors design. The necessary next step is to save a mdmt "template file." However in order to do that, you have to enter the name of the mdm file you want to save! You enter *radon.mdm* in the appropriate edit field, then click on *Save mdmt file.*



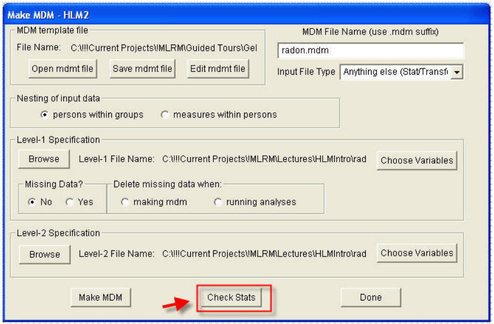
Constructing the MDM File

Once you've done saved the mdmt file, you can make the MDM file by clicking on the *make MDM* file at the bottom left.



Constructing the MDM File

At this point, you are strongly advised to Examine the basic statistics for the MDM file you have just created. You do this by clicking on the *Check Stats* button as shown below.



Constructing the MDM File

Examine the statistics, see if you have chosen the correct variables, and check whether the descriptive statistics make sense.

Then click *Done*

HLM2MDM.STS - Notepad

File Edit Format View Help

LEVEL-1 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
RADON	919	1.22	0.85	-2.30	3.88
FLOOR	919	0.17	0.37	0.00	1.00

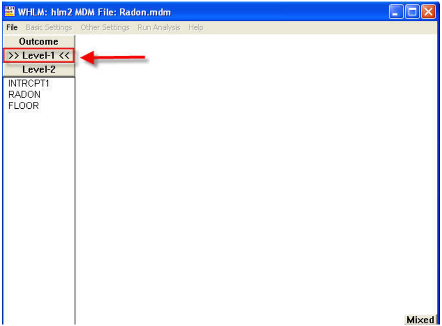
LEVEL-2 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
URANIUM	85	0.01	0.38	-0.88	0.53

Level-1 Specification

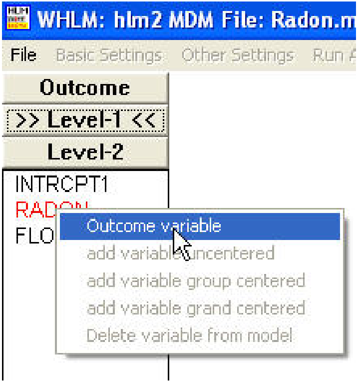
The next step is to specify the model.

Because this model is so fundamental, there isn't much specifying to do. Note that the Level-1 Button is highlighted.



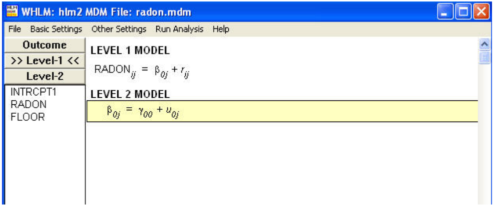
Level-1 Specification

our first step is to choose the Level-1 outcome variable. Click on *RADON* and a flying menu will open. Choose *Outcome variable*. You have now selected your outcome variable.



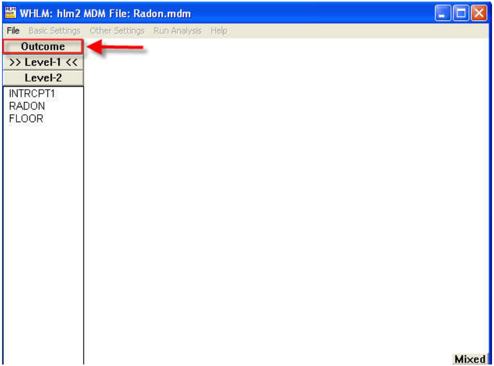
Constructing the MDM File

Next, you will see a window open up. This window will contain the current model specification in HLM notation. Note that a baseline Level-2 model has already been specified. Normally, you would next enter the Level-2 specification, but in this case, we are actually finished.



Outcome Variable Specifications

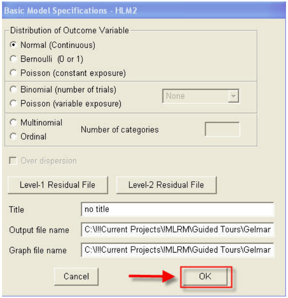
The next step is to specify the characteristics of the outcome variable. Click on the *Outcome* button.



Outcome Variable Specification

You'll see the dialog pictured below. Because we're assuming a normally distributed outcome, you don't have to do anything, although, if you wish, you could save residual files for analysis by another program.

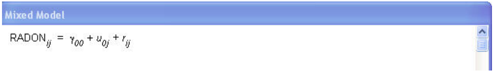
Just click *Ok*. Note: if you don't do this, your model will not be specified! Most modern software assumes a default (in this case a normal outcome variable) but HLM does not.



Constructing the MDM File

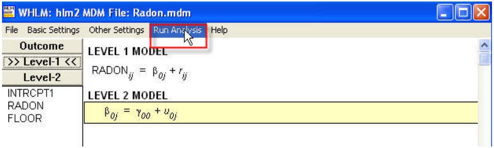
If you wish, HLM will automatically combine the two models into a single *mixed model*, which might be especially useful if you wish to use another program (like R) to analyze the model. Simply click on the *Mixed* button in the lower right corner of the main window.

In this case we can see that this model is indeed simply the 1-Way random-effects ANOVA.



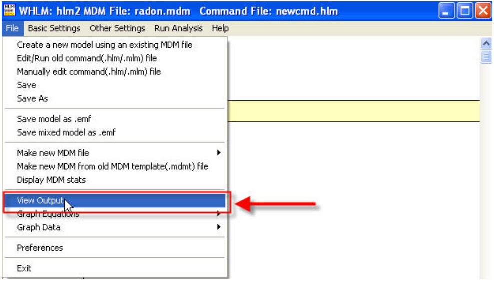
Analyzing the Model

Click on the *Run Analysis* button. HLM will open a DOS window and the model will run. There will be a brief pause near the end of calculations. Don't interrupt! The window will eventually shut.



Viewing Output

The output window will not open automatically. You need to select the *File–>View Output* menu option.



Viewing Output

The text window has a lot of superfluous information in it.

Halfway down, we encounter the results of estimation. I've excerpted key results.

```
The value of the likelihood function at iteration 6 = -1.129721E+003
The outcome variable is    RADON

 Final estimation of fixed effects:
 ----------------------------------------------------
                                   Standard             Approx.
    Fixed Effect         Coefficient  Error      T-ratio   d.f.     P-value
 ----------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00           1.312564   0.048894   26.845        84     0.000
 ----------------------------------------------------

 Final estimation of variance components:
 ----------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 ----------------------------------------------------
 INTRCPT1,       U0        0.30943       0.09574    84     226.17987    0.000
  level-1,       R         0.79789       0.63663
 ----------------------------------------------------

 Statistics for current covariance components model
 --------------------------------
 Deviance                       = 2259.442314
 Number of estimated parameters = 2
```

Comparing HLM and R Basic Output

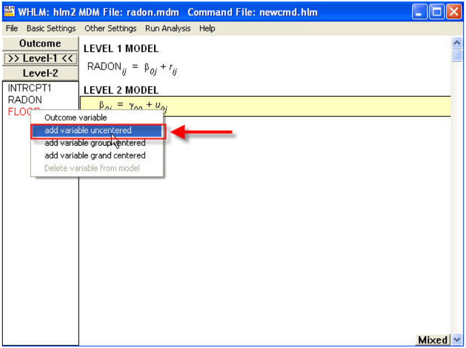
Let's compare these results to comparable results in R. We can see that they are essentially the same, although HLM includes a significance test that is not reported by R.

```
Linear mixed model fit by REML
Formula: radon ~ 1 + (1 | county)
  AIC  BIC logLik deviance REMLdev
 2265 2280  -1130     2255    2259
Random effects:
 Groups   Name        Variance Std.Dev.
 county   (Intercept) 0.095813 0.30954
 Residual             0.636621 0.79789
Number of obs: 919, groups: county, 85

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1.31257    0.04891   26.84
```

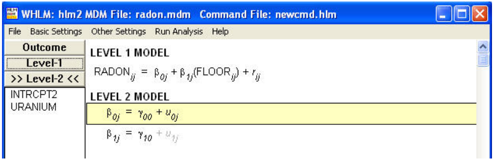
Adding a Floor Predictor

Now we wish to add `floor` as a predictor at level 1. In this case, we do not center the variable. Click on the variable name as shown in the snapshot below.



Checking the Model

You'll notice the model has changed. Notice that the random component for the slope at level two is greyed out. You can toggle the random components on and off by clicking on them.



Analyzing the Output

Next, we analyze the data, after saving our model with an appropriate name. Again, I have excerpted key results.

```
The outcome variable is    RADON
Final estimation of fixed effects:
----------------------------------------------------
                                      Standard           Approx.
   Fixed Effect          Coefficient  Error      T-ratio   d.f.     P-value
----------------------------------------------------
For       INTRCPT1, B0
   INTRCPT2, G00            1.461579  0.051564    28.345        84     0.000
For    FLOOR slope, B1
   INTRCPT2, G10           -0.692979  0.070430    -9.839       917     0.000
----------------------------------------------------
Final estimation of variance components:
----------------------------------------------------
Random Effect            Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
----------------------------------------------------
INTRCPT1,       U0         0.32812       0.10766    84     277.04086    0.000
 level-1,       R          0.75560       0.57093
----------------------------------------------------


Statistics for current covariance components model
---------------------------------
Deviance                          = 2169.467425
Number of estimated parameters = 2
```

Comparing with R Output

Previously, we saw that the R output from `lmer()` produced these virtually identical results.

```
> fit.1 ← lmer(radon ~ floor + (1 | county))
> summary(fit.1)
```

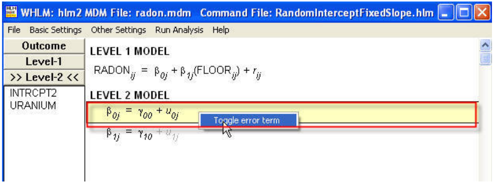
```
Linear mixed model fit by REML
Formula: radon ~ floor + (1 | county)
  AIC  BIC logLik deviance REMLdev
 2179 2199  -1086     2164    2171
Random effects:
 Groups   Name        Variance Std.Dev.
 county   (Intercept) 0.108    0.328
 Residual             0.571    0.756
Number of obs: 919, groups: county, 85

Fixed effects:
            Estimate Std. Error t value
(Intercept)   1.4616     0.0516   28.34
floor        -0.6930     0.0704   -9.84

Correlation of Fixed Effects:
      (Intr)
floor -0.288
```

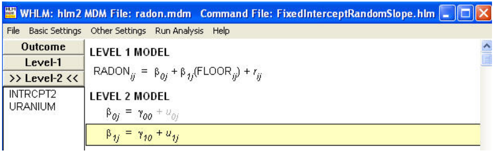
Specifying the Model

Shifting to a model where the intercept is fixed across groups but the slope for a `floor` predictor varies is very simple. We simply disconnect one error term and connect the other at level-2. First, point to the first line of the level-2 model and right click, opening a window to *Toggle the error term* as shown below. Then click to grey the error term. This means the u_{0j} term will not be included in the model.



Specifying the Model

Next, right click the second line of the model and toggle *on* the u_{1j} term. Now the model is set up to have varying slopes but a fixed intercept. When you are done, your model should look like this:



Analyzing the Model

Analyze the model, and examine the data. Here are some of the results.

```
The outcome variable is    RADON

 Final estimation of fixed effects:
 ----------------------------------------------------
                                       Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00           1.326744   0.029321    45.248       917    0.000
 For    FLOOR slope, B1
    INTRCPT2, G10          -0.554486   0.089261    -6.212        84    0.000
 ----------------------------------------------------

 Final estimation of variance components:
 ----------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 ----------------------------------------------------
    FLOOR,       U1        0.34040       0.11587    59      87.04679    0.010
  level-1,       R         0.81152       0.65856
 ----------------------------------------------------
 Note: The chi-square statistics reported above are based on only 60 of 85
units that had sufficient data for computation.  Fixed effects and variance
components are based on all the data.
```

Comparing Results with R

Compare this result with the corresponding R output.

```
> fit.2 ← lmer(radon ~ floor + (floor-1|county))
> summary(fit.2)
```

```
Linear mixed model fit by REML
Formula: radon ~ floor + (floor - 1 | county)
  AIC  BIC logLik deviance REMLdev
 2259 2278  -1125     2242    2251
Random effects:
 Groups   Name  Variance Std.Dev.
 county   floor 0.115    0.340
 Residual       0.659    0.812
Number of obs: 919, groups: county, 85

Fixed effects:
            Estimate Std. Error t value
(Intercept)   1.3267     0.0293    45.2
floor        -0.5546     0.0892    -6.2

Correlation of Fixed Effects:
      (Intr)
floor -0.329
```

Specifying the Model

Simply toggle on the error term at both levels. By now this should be a snap. To double check, examine the full mixed model and see how it compares to this:

$$\text{RADON}_{ij} = \gamma_{00} + \gamma_{10}\text{FLOOR}_{ij} + u_{0j} + u_{1j}\text{FLOOR}_{ij} + r_{ij} \quad (7)$$

Then analyze the model.

Analyzing the Output

Your basic output should look like this:

```
The value of the likelihood function at iteration 81 = -1.083243E+003
 The outcome variable is    RADON

 Final estimation of fixed effects:
 ----------------------------------------------------
                                       Standard            Approx.
    Fixed Effect           Coefficient   Error      T-ratio    d.f.     P-value
 ----------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00           1.462763   0.053874    27.152        84     0.000
 For    FLOOR slope, B1
    INTRCPT2, G10          -0.680984   0.087669    -7.768        84     0.000
 ----------------------------------------------------
 Final estimation of variance components:
 ----------------------------------------------------
 Random Effect             Standard      Variance     df     Chi-square  P-value
                           Deviation     Component
 ----------------------------------------------------
 INTRCPT1,        U0         0.34875       0.12163     59     231.70805    0.000
    FLOOR slope, U1          0.34469       0.11881     59      81.23290    0.029
  level-1,        R          0.74612       0.55670
 ----------------------------------------------------


Note: The chi-square statistics reported above are based on only 60 of 85
units that had sufficient data for computation.  Fixed effects and variance
components are based on all the data.
```

Comparing to R Output

```
> fit.3 ← lmer(formula = radon ~ floor + (1 + floor | county))
> summary(fit.3)
```

```
Linear mixed model fit by REML
Formula: radon ~ floor + (1 + floor | county)
  AIC  BIC logLik deviance REMLdev
 2180 2209  -1084     2161    2168
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 county   (Intercept) 0.122    0.349
          floor       0.118    0.344    -0.337
 Residual             0.557    0.746
Number of obs: 919, groups: county, 85

Fixed effects:
            Estimate Std. Error t value
(Intercept)   1.4628     0.0539   27.15
floor        -0.6811     0.0876   -7.78

Correlation of Fixed Effects:
      (Intr)
floor -0.381
```

Adding Predictors at Level 2

The model can be enhanced by adding soil uranium as a predictor at level 2. In this case, we predict both the level-1 slopes and the level-1 intercepts from uranium. In HLM notation, the models become, at level 1,

$$\text{RADON}_{ij} = \beta_{0j} + \beta_{1j}\text{FLOOR}_{ij} + r_{ij} \tag{8}$$

and, at level 2,

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}\text{URANIUM}_j + u_{0j} && (9) \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}\text{URANIUM}_j + u_{1j} && (10) \end{aligned}$$

Setting Up the Model in HLM

By now, specifying the model in HLM should be a breeze for you. Simply click on the *Level-2* button, then click on each line of the level 2 model and add the `uranium` variable to it.

When you are done, your model should match that shown on the preceding slide.

Analyzing Model Results

Analyze the model, and you should see results like these:

```
 The outcome variable is    RADON

 Final estimation of fixed effects:
 ----------------------------------------------------
                                       Standard             Approx.
    Fixed Effect         Coefficient   Error      T-ratio   d.f.     P-value
 ----------------------------------------------------
 For       INTRCPT1, B0
    INTRCPT2, G00           1.468574   0.035310    41.591        83     0.000
     URANIUM, G01           0.808049   0.090780     8.901        83     0.000
 For    FLOOR slope, B1
    INTRCPT2, G10          -0.670744   0.084647    -7.924        83     0.000
     URANIUM, G11          -0.418380   0.227729    -1.837        83     0.069
 ----------------------------------------------------
 Final estimation of variance components:
 ----------------------------------------------------
 Random Effect           Standard      Variance     df    Chi-square  P-value
                         Deviation     Component
 ----------------------------------------------------
 INTRCPT1,       U0        0.12509       0.01565    58      87.02121    0.008
    FLOOR slope, U1        0.30977       0.09596    58      77.51349    0.044
  level-1,       R         0.74934       0.56151
 ----------------------------------------------------


Note: The chi-square statistics reported above are based on only 60 of 85
units that had sufficient data for computation.  Fixed effects and variance
components are based on all the data.


 Statistics for current covariance components model
 ------------------------------
 Deviance                       = 2124.742483
 Number of estimated parameters = 4
```

Interpreting the Output

How does the level-1 model react to changes in uranium level across counties? Let's examine the printed output and try to make a few predictions.

- Suppose soil uranium level is at the 50th percentile among counties. What would you expect the line relating floor to radon level to look like?
- How about the 87.5th percentile? The 12.5th percentile?

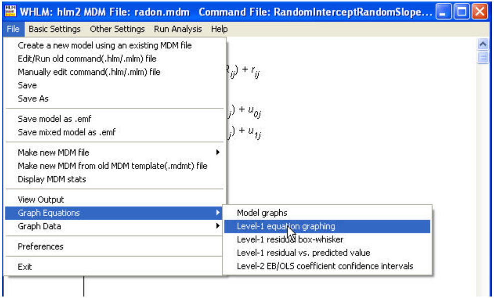
Interpreting the Output

How does the level-1 model react to changes in uranium level across counties? Let's examine the printed output and try to make a few predictions.

- Suppose soil uranium level is at the 50th percentile among counties. What would you expect the line relating floor to radon level to look like?
- How about the 87.5th percentile? The 12.5th percentile?

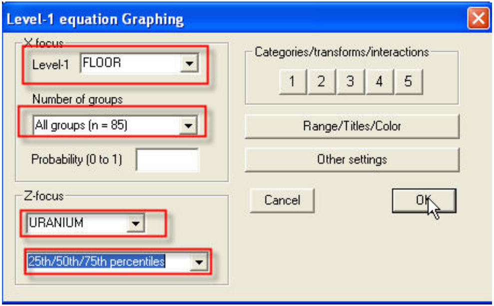
Plotting the Regression Lines in HLM

Let's use HLM to plot the regression. Go to the *Graph Equations –> Level-1 equation graphing* menu, as shown below.



Plotting the Regression Lines in HLM

Set the menu items as shown below, and click OK.



Plotting the Regression Lines in HLM