Introducing the Multilevel Model for Change

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

GCM, 2010

Introduction

In this lecture, we introduce the *multilevel model for change*. We begin with a quick review of multilevel models, and then specialize to the multilevel growth curve model.

Hierarchical Data Structures

In many situations in data analysis, the data are *nested*, which often implies a *hierarchical* structure. For example:

- soldiers are nested within platoons, which are nested within companies, etc.
- Students are nested within classes, which are nested within schools, which are nested within school districts, etc.
- Of special interest in our current application is how, in longitudinal data analysis, *observations can be nested within individuals*.

A Not-So-Simple "Simple Example"

In his introductory text, Goldstein (1999) considers a sample of students from the Junior School Project (JSP) data.

- There were over 1000 students nested within 50 schools
- Each student took the Raven progressive matrices test (an ability measure) and also was assessed several times in English and Mathematics achievement.
- Suppose we consider, for the time being, just the relationship between the Raven test and the first test result in Mathematics.

Notational Choices

There are many ways we can write a multilevel model. We'll look at several here. We'll start with a classic two-level approach used by Raudenbush and Bryk. Keep in mind that you have to process the *ideas*. The notational variations (use of different letters, etc.) that you will see in various books and articles are virtually endless.

The Level-1 Model

At Level 1, we can analyze the relationship between the Raven test and Math Achievement within each school. In this case, if we fit a simple linear regression line to the data, the model can be written for the ith student in the jth school as

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \tag{1}$$

The r_{ij} are regression residuals and are assumed to be independently and identically normally distributed with a mean of 0 and a variance of σ^2.

The Level-2 Model

Notice how each school has its own slope and intercept. The slope and intercept define the regression line. These slopes and intercepts can be viewed as arising through random normal variation around central value. Specifically, the Level-2 model is

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (3)$$

The β_0s and β_1s are allowed to correlate, so they have a covariance matrix $\boldsymbol{T}$ that is not necessarily diagonal.

The Combined Mixed Model

By simply substituting the Level-2 model into the Level-1 expression, we get the combined (mixed) model which contains both fixed (the γs) and random (the βs) effects. We get

$$\begin{aligned} Y_{ij} &= \gamma_{00} + u_{0j} + (\gamma_{10} + u_{1j})X_{ij} + r_{ij} \\ &= \gamma_{00} + \gamma_{10}X_{ij} + (u_{0j} + u_{1j}X_{ij} + r_{ij}) \qquad (4) \\ &= \gamma_{00} + \gamma_{10}X_{ij} + r^*_{ij} \qquad (5) \\ &= \gamma_{00} + \gamma_{10}X_{ij} + (u_{0j} + u_{1j}X_{ij}) + r_{ij} \qquad (6) \end{aligned}$$

The above *multilevel* model is a special case of the *linear mixed effects model*. The γs are *fixed effects*, and the us are *random effects*, i.e., random variables having a distribution.

The Combined Model – Matrix Notation

As I already mentioned, we can express the above model in a wide variety of notations. Here is one matrix notation. First, let's simply convert the scalar notation directly into matrix notation using the same lettering. Let $\boldsymbol{y}_j$ represent the vector of scores of the individuals in group j. Then

$$\boldsymbol{y}_j = \boldsymbol{X}_j\boldsymbol{\gamma} + \boldsymbol{X}_j\boldsymbol{u}_j + \boldsymbol{r}_j \tag{7}$$

Note that, in the above, the scores in $\boldsymbol{X}_j$ can be either design codes or actual scores on observed variables, because of the longstanding tradition in regression modeling of treating observed scores on predictors as *fixed* regardless of whether they are actually fixed scores or random variables.

The Combined Model – Matrix Notation

Suppose we have just two schools and there are 3 students in school 1 and 4 students in school 2. The notation corresponds to the data as follows:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \boldsymbol{\gamma} + \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \end{bmatrix}$$

$$\left[\begin{array}{c} Y_{11} \\ Y_{21} \\ Y_{31} \\ \hline Y_{12} \\ Y_{22} \\ Y_{32} \\ Y_{42} \end{array}\right] = \left[\begin{array}{cc} 1 & X_{11} \\ 1 & X_{21} \\ 1 & X_{31} \\ \hline 1 & X_{12} \\ 1 & X_{22} \\ 1 & X_{32} \\ 1 & X_{42} \end{array}\right] \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} + \left[\begin{array}{cccc} 1 & X_{11} & 0 & 0 \\ 1 & X_{21} & 0 & 0 \\ 1 & X_{31} & 0 & 0 \\ \hline 0 & 0 & 1 & X_{12} \\ 0 & 0 & 1 & X_{22} \\ 0 & 0 & 1 & X_{32} \\ 0 & 0 & 1 & X_{42} \end{array}\right] \begin{bmatrix} u_{01} \\ u_{11} \\ u_{02} \\ u_{12} \end{bmatrix} + \left[\begin{array}{c} r_{11} \\ r_{21} \\ r_{31} \\ \hline r_{12} \\ r_{22} \\ r_{32} \\ r_{42} \end{array}\right]$$

The General Linear Mixed-Effects Model

The preceding can be seen (with some minor changes in lettering) to be a special case of the General Linear Mixed-Effects Model. Pinheiro and Bates (2000) in their well-known book, *Mixed-Effects Models in S and S-Plus* use the following notation for the GLMM for a single level of grouping, i.e., what Raudenbush and Bryk refer to as a two-level model.

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i \tag{8}$$

$$\boldsymbol{b}_1 \sim N(\mathbf{0}, \boldsymbol{\Psi}), \quad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2\boldsymbol{\Lambda}_i) \tag{9}$$

where $\boldsymbol{X}_i$ is the fixed effects regressor matrix for the ith unit, and $\boldsymbol{Z}_i$ is the random effects regressor matrix, which usually contains a subset (perhaps all) of the columns of $\boldsymbol{X}_i$. The vector $\boldsymbol{\beta}$ contains fixed effects, while $\boldsymbol{b}_i$ contains the random effects.

Fitting the Two-Level Model

The two level model with random slopes and intercepts can be fit with two rather different programs.

- One program, R, is free, and very general.
- The other, HLM, is a well-known commercial program designed to fit hierarchical linear models.
- R requires that you compute the full mixed model, as shown in Equation 6.
- HLM allows you to input the Level-1 and Level-2 models separately, and declare whether the effects for each are fixed or random.

To get a feel for how they work, we'll try both programs on the high school data.

Fitting the Model with R

To fit LME models in R, we need to use the `lmer` function in the `lme4` package. We begin by loading in the data file, which is in long format. We'll convert it to wide format and remove missing data.

```
> library(foreign)
> jsp.long <- read.table("jsp.csv", header=TRUE,
+ sep = ",")
> jsp.wide <- reshape(data=jsp.long,direction="wide",
+ timevar="year",idvar="pupil")
> jsp.wide <- na.omit(jsp.wide)
> write.table(jsp.wide,"jsp.wide.txt",row.names=F,
+ col.names=T,sep=",")
```

Fitting the Model with R

Then, we enter the model, using the appropriate syntax.

```
> full.model <- lmer(maths.0 ~ ravens.0 +
+ (1 + ravens.0 | school.0),data=jsp.wide)
> full.model

Linear mixed model fit by REML
Formula: maths.0 ~ ravens.0 + (1 + ravens.0 | school.0)
   Data: jsp.wide
  AIC  BIC logLik deviance REMLdev
 5618 5647  -2803     5601    5606
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 school.0 (Intercept) 19.6206  4.430
          ravens.0     0.0277  0.166    -0.935
 Residual             30.2870  5.503
Number of obs: 887, groups: school.0, 48

Fixed effects:
            Estimate Std. Error t value
(Intercept)   7.9306     1.1241    7.06
ravens.0      0.6917     0.0428   16.14

Correlation of Fixed Effects:
         (Intr)
ravens.0 -0.962
```

Fitting the Model with HLM

With HLM, rather than fitting the mixed model directly, we input the Level-1 and Level-2 models separately. This is a rather straightforward (if somewhat idiosyncratic) process.

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 model that we examined, and set it up and analyze it in HLM.

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

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 (if you have the Student Version), or comma-delimited ASCII files with a header containing column names (if you have the full version). 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

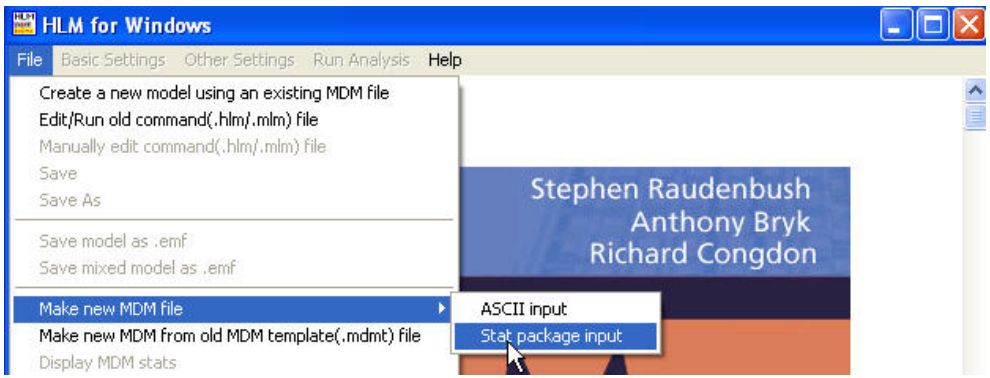
The link between the level-1 and level-2 models in the HLM parameterization is the subscript j, which refers in the current example to the `school.0` variable. To set up the data for HLM, we need two files, one for the level-1 variables, one for the level-2 variables. Each file has to be sorted in ascending order of the ID variable.

We need to include `ravens.0`, `math.0`, and `school.0` in the level-1 file, and `school.0` in the level-2 file. Since all the variables we need are in one file, we are ready to go.

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 the file `jsp.wide.txt` available, or, if you are using the Student Version, the SPSS file `jsp.wide.sav`. We will show the procedure for the SPSS file. Begin by starting up HLM. Then click on the *Make New MDM File -> Stat Package Input* menu option, whether you are loading an SPSS file or a comma-delimited text file! (This is counterintuitive and very poor human factors design. One would expect this to be found under the ASCII file input node.)

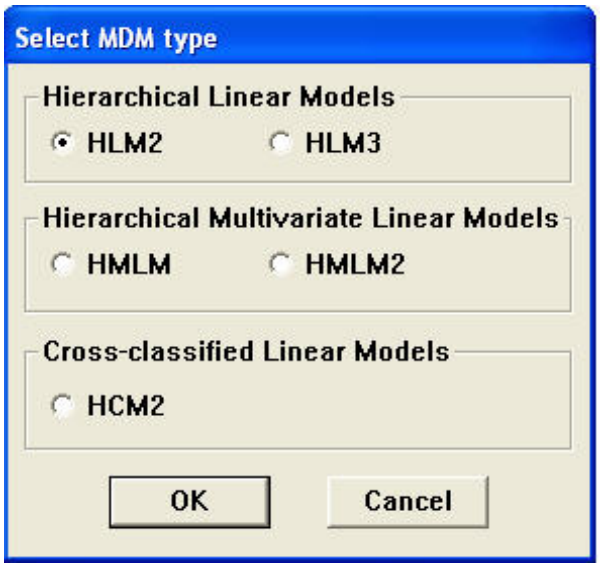


Constructing the MDM File

Next, you will be asked to select a program.

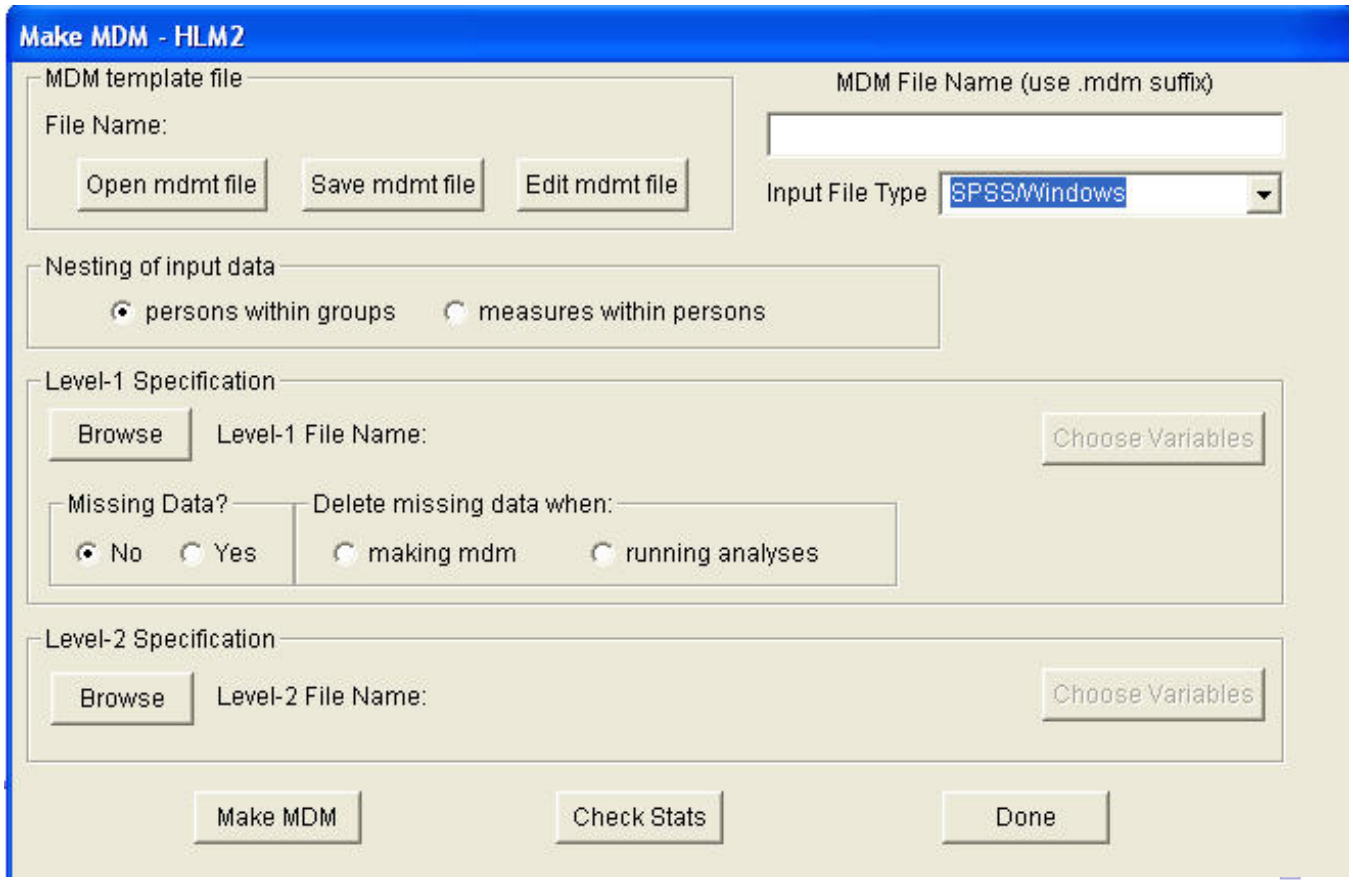
Select HLM2.

Then click on the OK button.



Constructing the MDM File

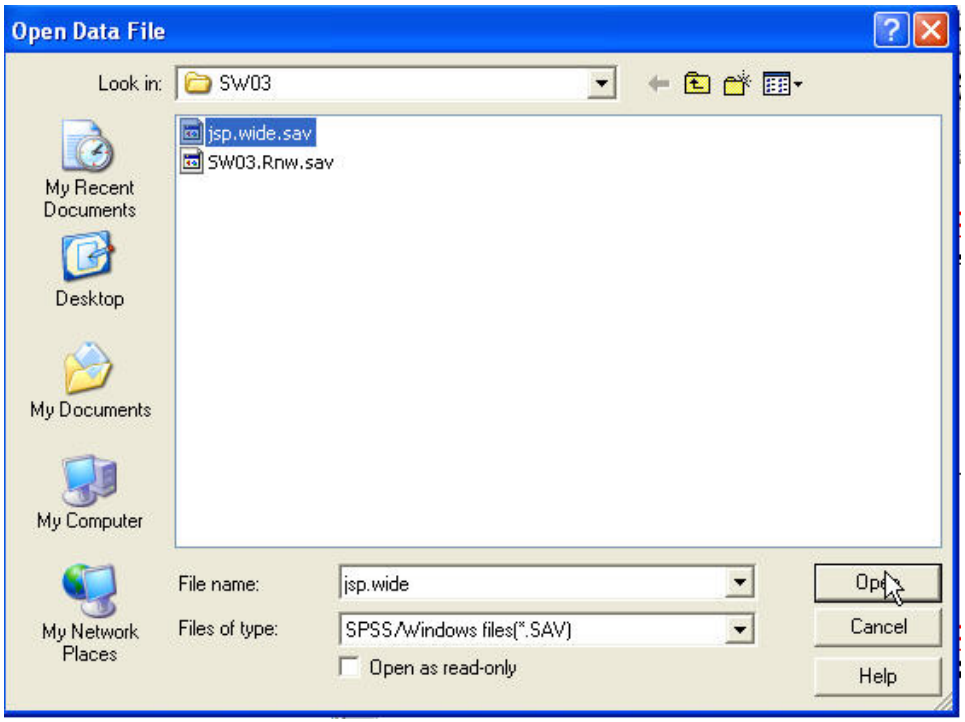
A large dialog box will open. Go to the drop-down list for file type, and select *SPSS/Windows*.



Constructing the MDM File

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

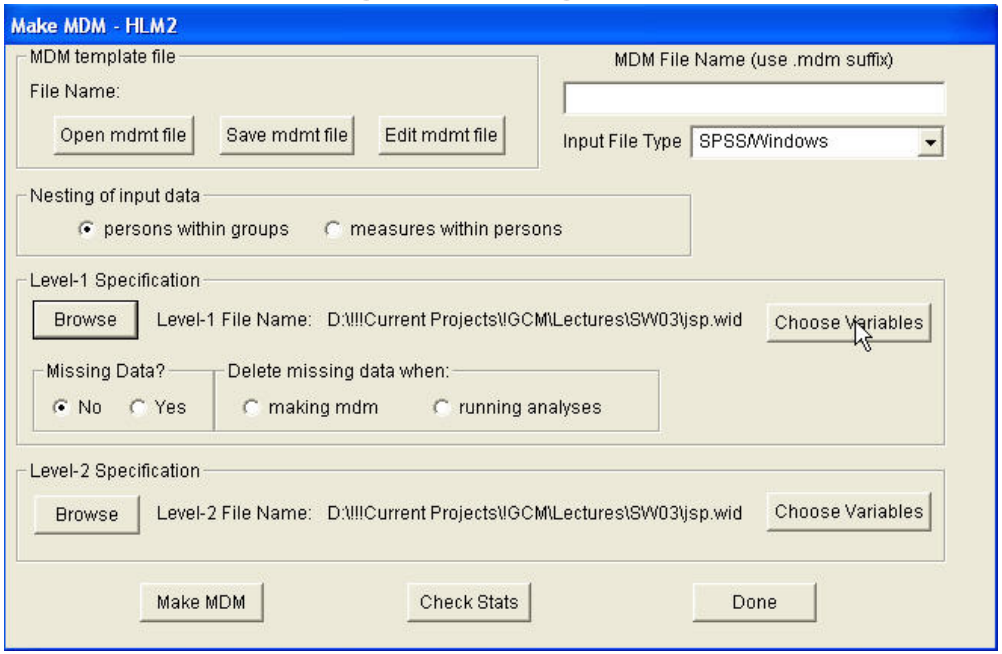
Select *jsp.wide.sav* and click on the *Open* button.



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.

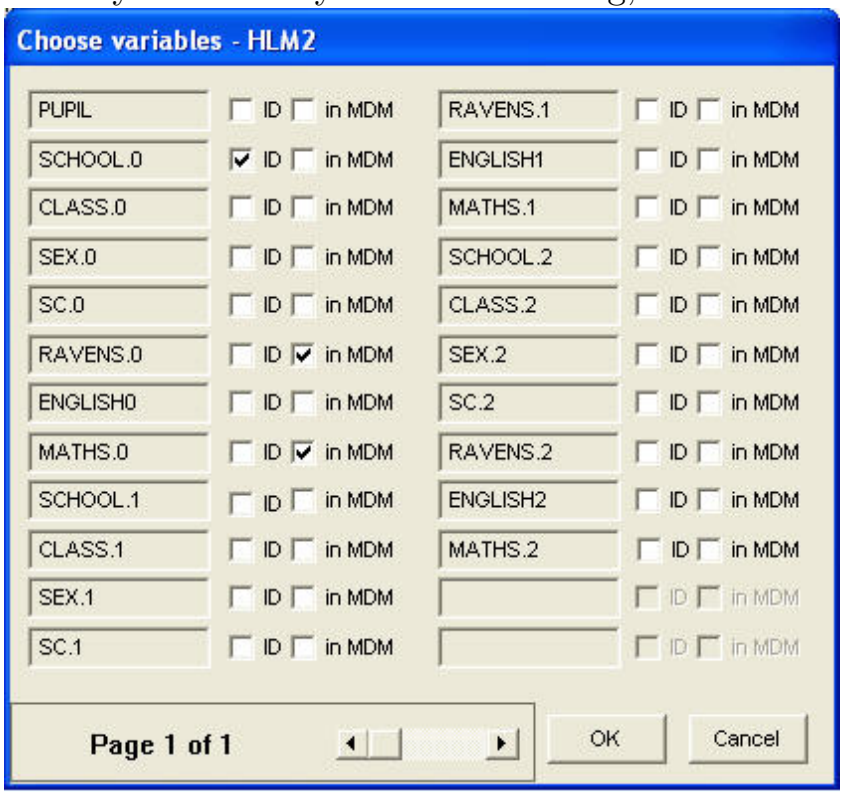
On the right side of that grouping is a button *Choose Variables*.



Click on it.

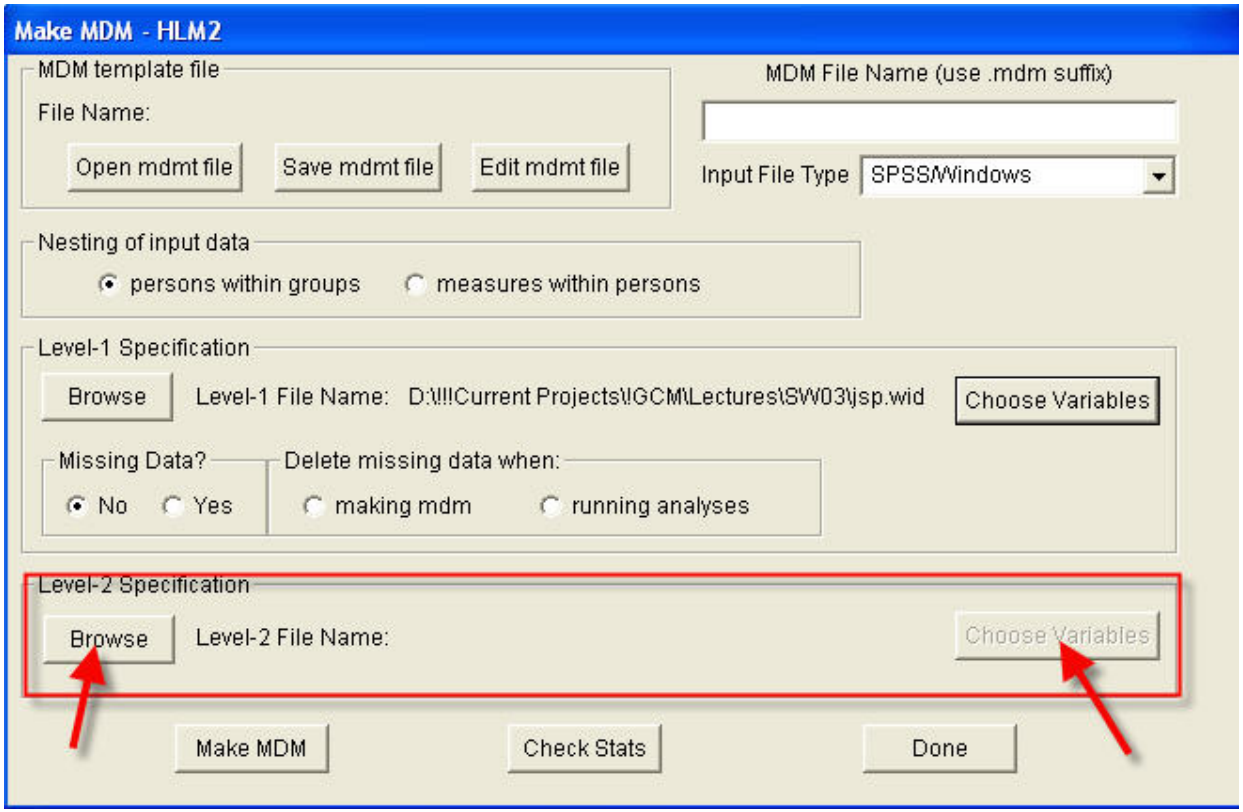
Constructing the MDM File

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 `SCHOOL.0`, and this variable is designated an ID variable. The variables `MATHS.0` and `RAVENS.0` 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. Again select the file *jsp.wide.sav*. Second, click on the *Choose Variables* button.

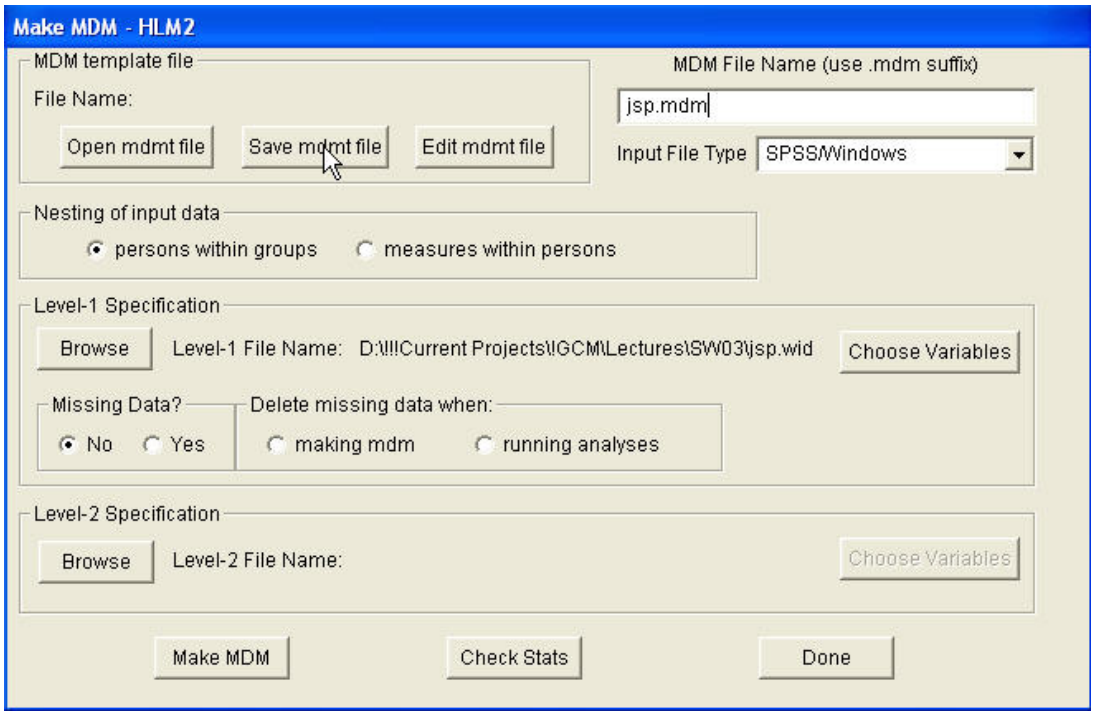


Constructing the MDM File

This will take you to another variable selection dialog. Again `school` is the ID variable, and we actually have no level-2 predictors in this model. However, HLM requires you to select at least one additional variable, so choose the same variables as for the level-1 file.

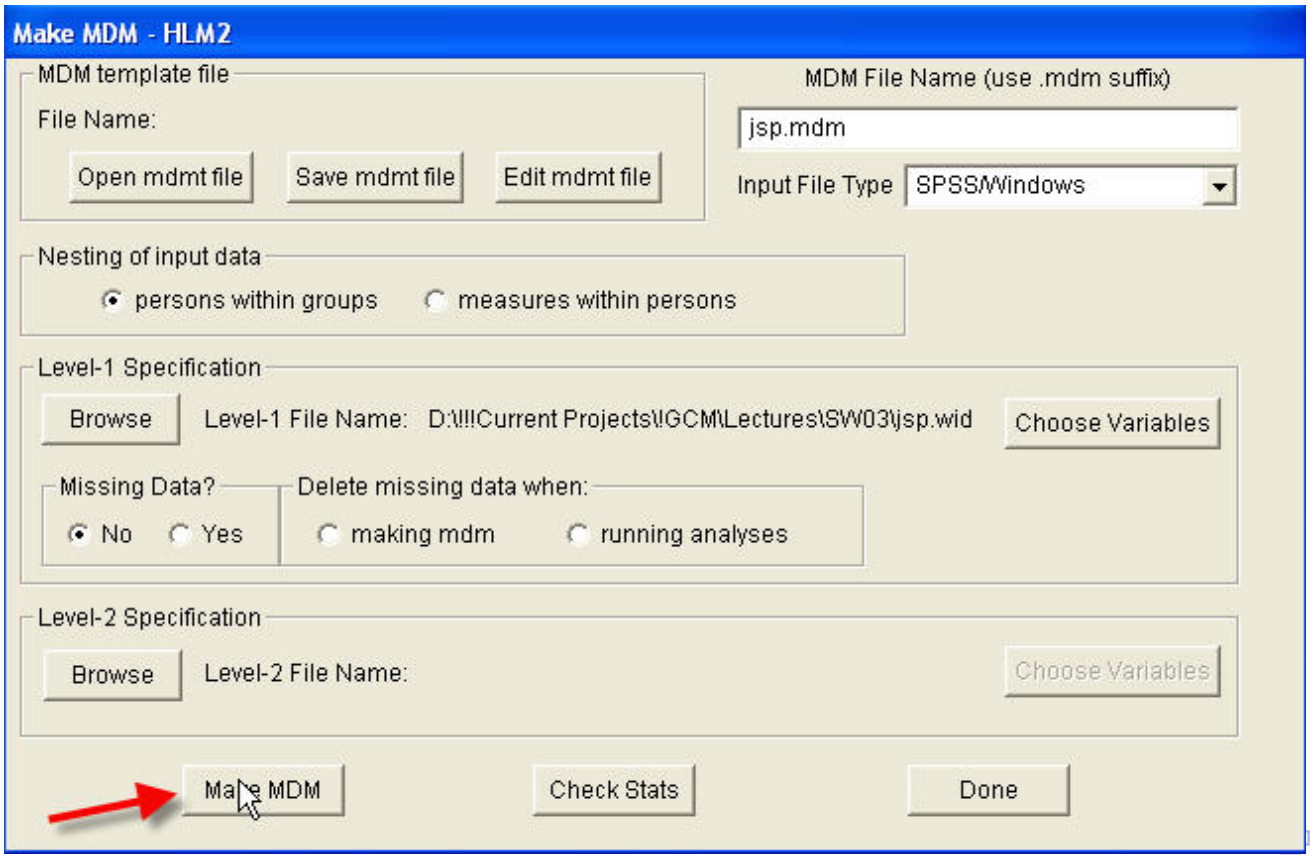
Constructing the MDM File

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 *jsp.mdm* in the appropriate edit field, then click on *Save mdmt file*.



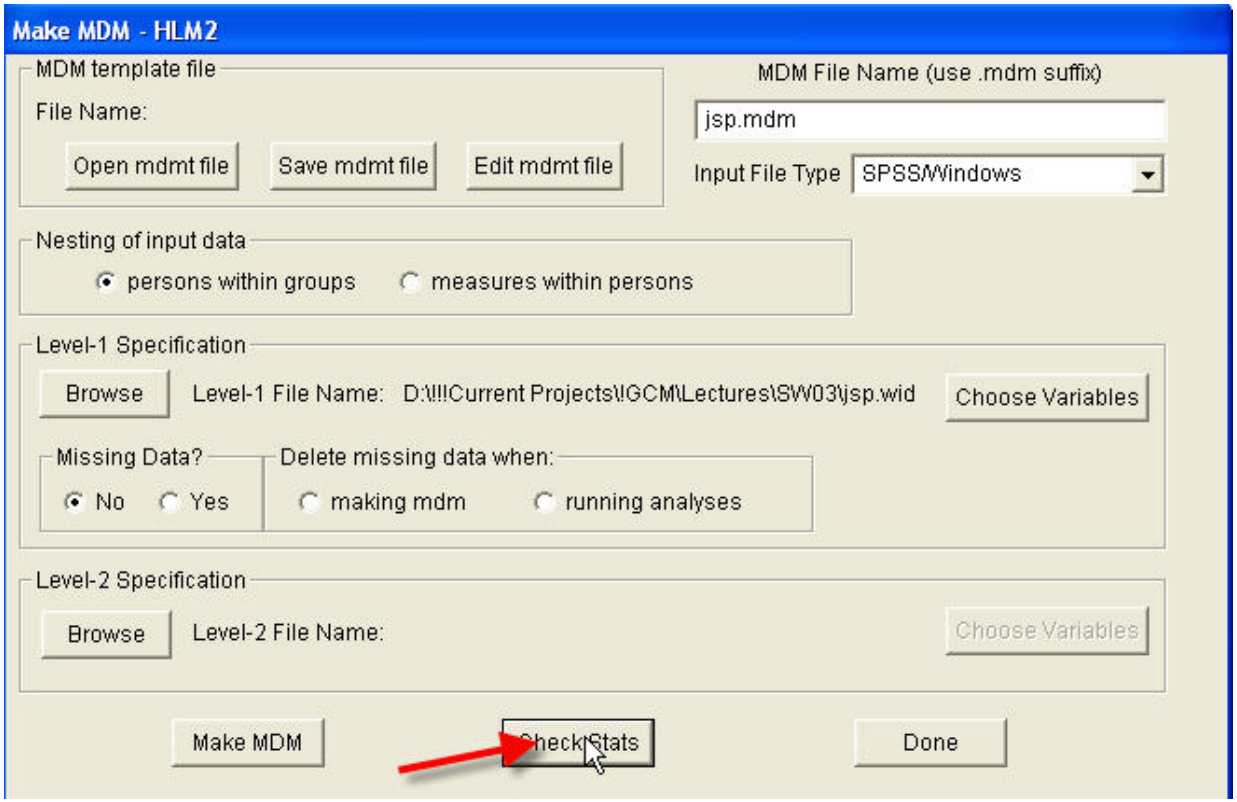
Constructing the MDM File

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
RAVENS.0	887	25.36	5.78	4.00	36.00
MATHS.0	887	25.67	6.99	5.00	40.00

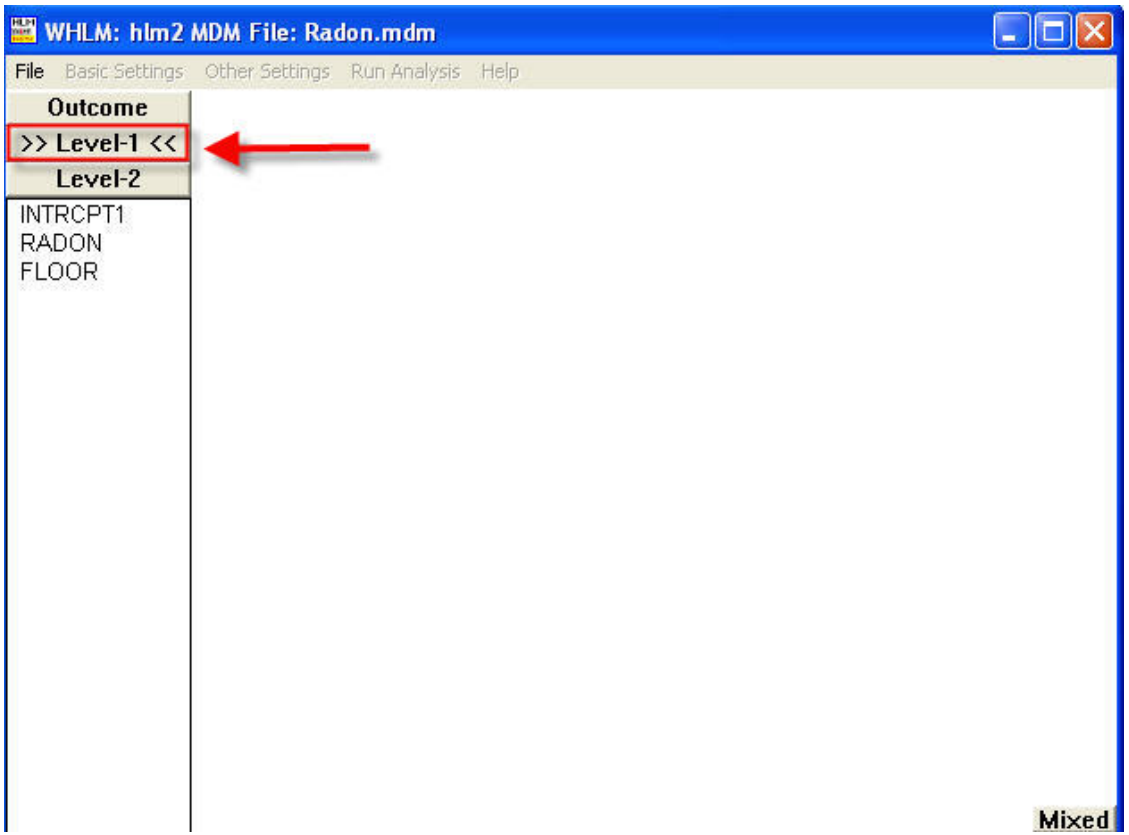
LEVEL-2 DESCRIPTIVE STATISTICS

VARIABLE NAME	N	MEAN	SD	MINIMUM	MAXIMUM
RAVENS.0	48	25.19	5.87	12.00	35.00
MATHS.0	48	25.48	7.46	12.00	40.00

Level-1 Specification

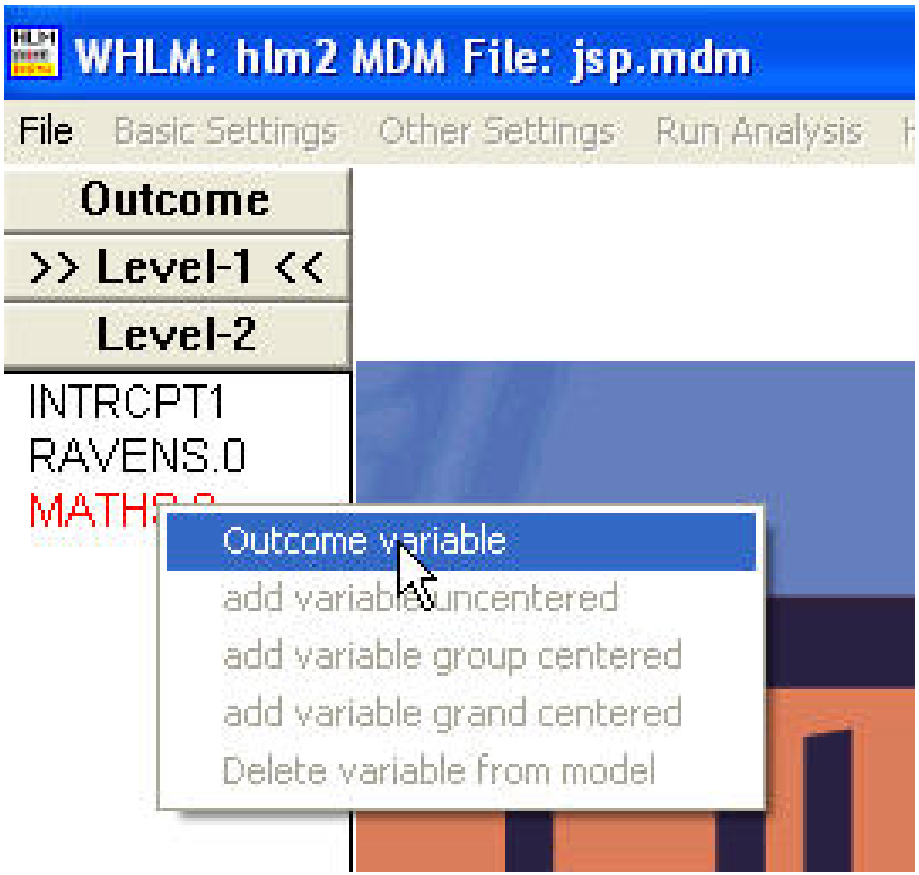
The next step is to specify the model.

Note that the Level-1 Button is highlighted.



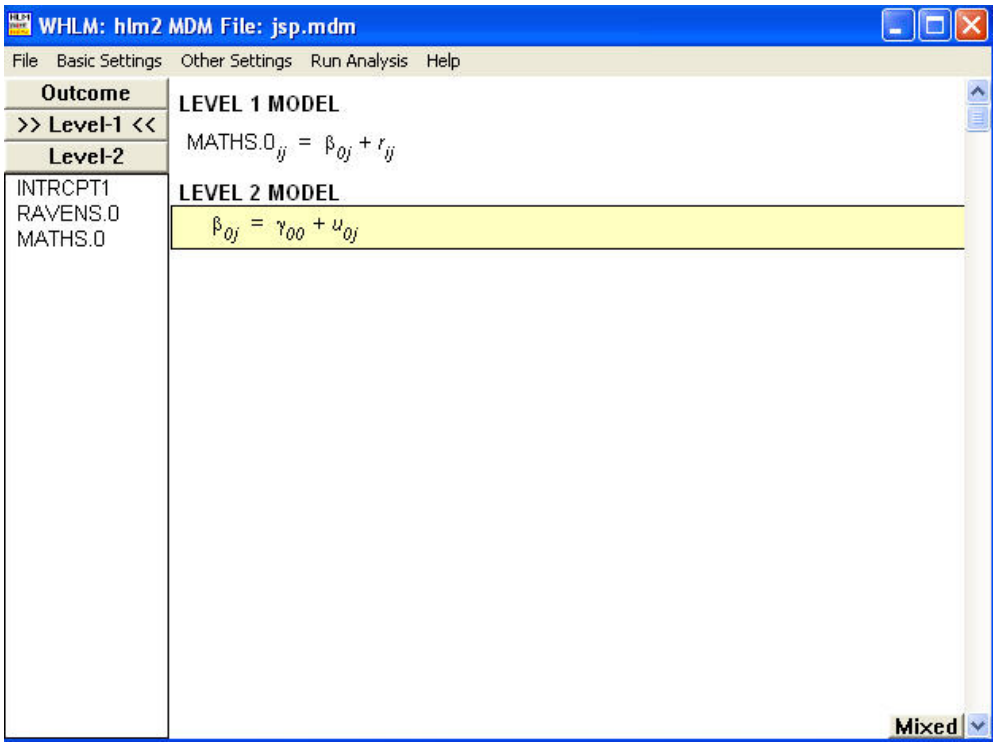
Level-1 Specification

our first step is to choose the Level-1 outcome variable. Click on *MATHS.0* and a flying menu will open. Choose *Outcome variable*. You have now selected your outcome variable.



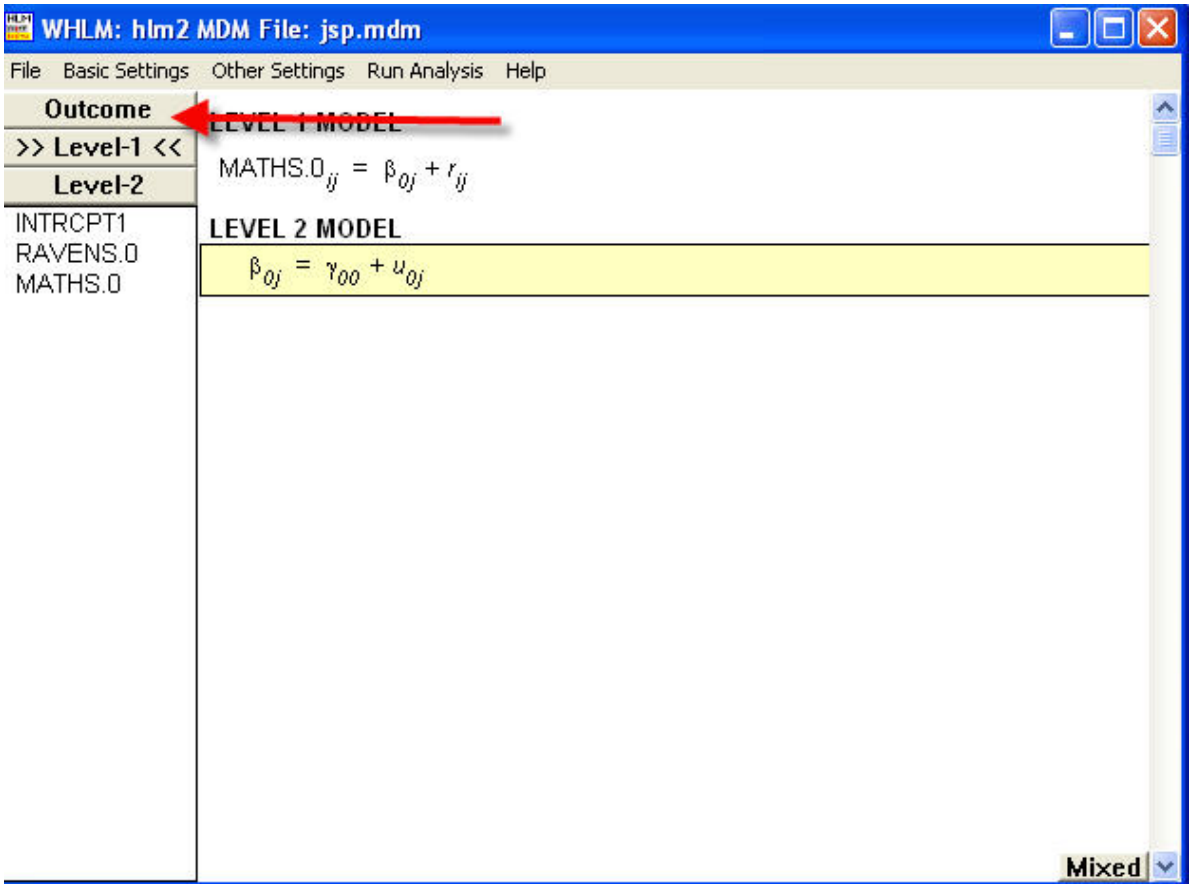
Constructing the MDM File

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

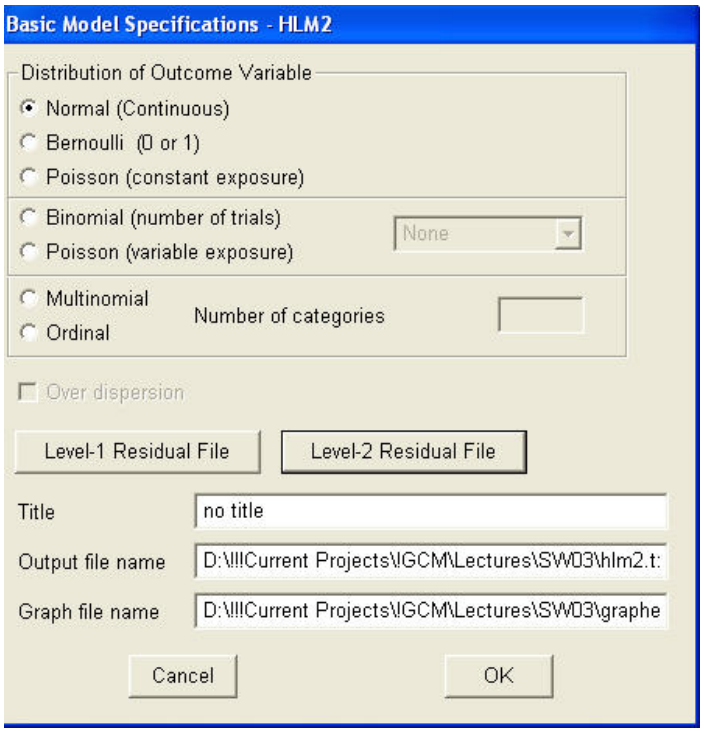
The next step is to specify the characteristics of the outcome variable. Click on the *Outcome* button.



Outcome Variable Specification

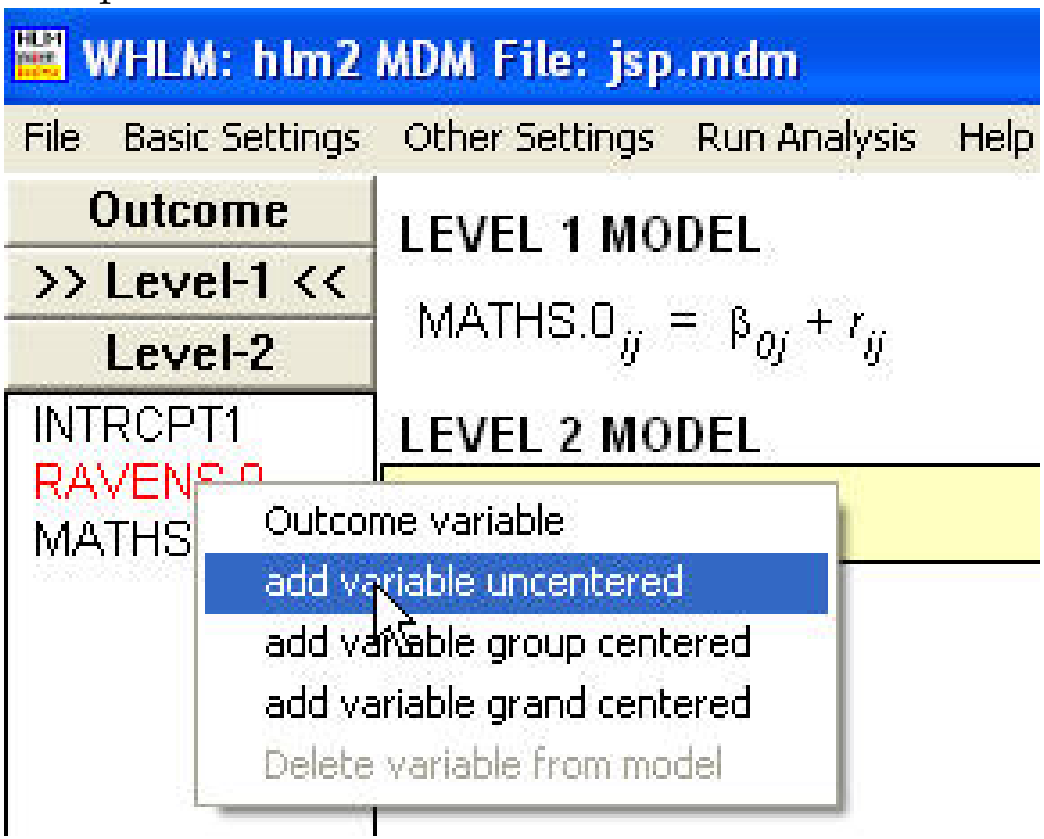
You'll see the dialog pictured below. Because we're assuming a normally distributed outcome, you don't have to do anything, although, if you wish, you could save residual files for analysis by another program.

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.



Adding a Level-1 Predictor

In this case, we wish to predict `MATHS.0` from `RAVENS.0` at level 1. So we click on the variable name and add it, uncentered. Notice how the model is changed automatically to reflect the new predictor.



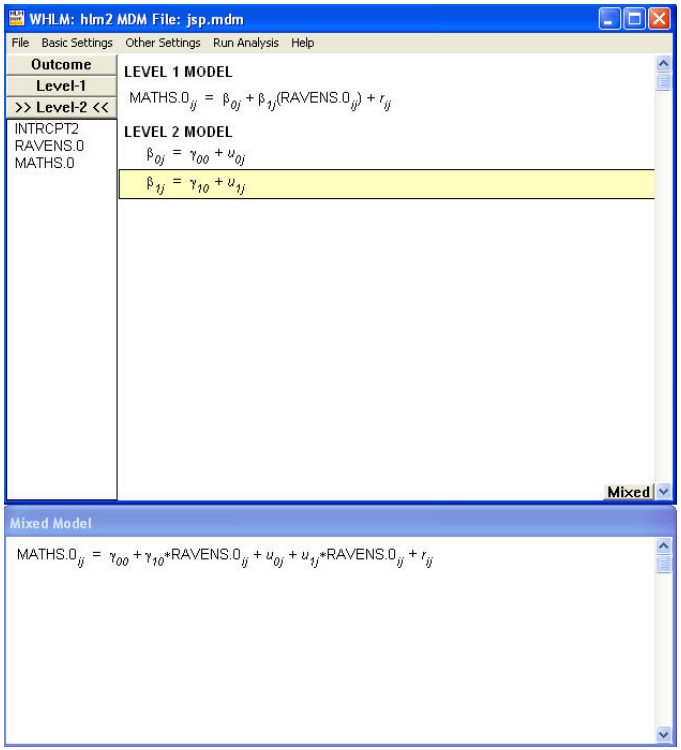
Adding a Level-2 Random Term

The level-2 model includes random slopes and intercepts. Notice how the random effect u_{1j} is greyed out in the level-2 model. Click on it and it is immediately added to the level-2 model.

Constructing the MDM File

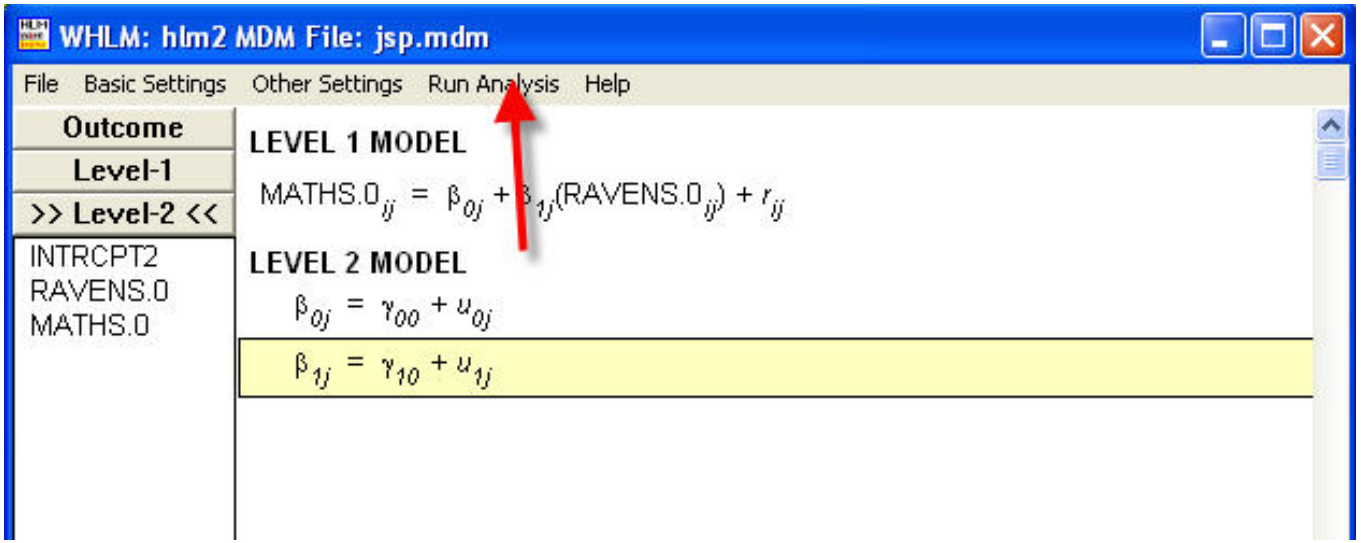
If you wish, HLM will automatically combine the two models into a single *mixed model*, which might be especially useful if you wish to use another program (like R) to analyze the model. Simply click on the *Mixed* button in the lower right corner of the main window.

In this case we can see that this model is indeed the same as what we had before.



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.