

Hierarchical Linear Modeling: A Primer¹ (Measures Within People)

R. C. Gardner
Department of Psychology

As noted previously, Hierarchical Linear Modeling (HLM) can be considered a particular instance of analysis of variance, and of course, analysis of variance can be viewed as a particular instance of multiple regression (Gardner, 2007). The three procedures test somewhat different null hypotheses but in the end they provide answers to comparable questions. It should come as no surprise, therefore, that when viewed from the point of view of “measures within people”, HLM can be shown to test hypotheses comparable to those tested by repeated measures analysis of variance. The purpose of this manuscript is to show the similarity between HLM (measures within people) and repeated measures analyses of variance, and to demonstrate how to use the computer package, HLM6. We will focus attention on three applications, (a) a single factor repeated measures analysis of variance, (b) an analysis with repeated measures on one factor and “Between Subjects” variability on the other, and (c) an analysis with repeated measures and a corresponding covariate. It can, of course, be extended to even more complex models.

Purpose and Rationale

In essence, the question being asked in the simplest form of the random-coefficients model when applied to measures within people (i.e., (a) as described above) is whether there is an association between the dependent variable and the repeated measures factor (i.e., a main effect of the repeated measures factor on the dependent variable). For (b), this is extended to determine whether the nature of the variation across the repeated measures factor is related to differences associated with the Between Subjects factor. For (c) it is akin to a repeated measures analysis of covariance with the covariate varying over the repeated measures. In analysis of variance terms, these are examined by means of main and interaction effects involving the repeated measures factor. In the context of HLM, these hypotheses are tested by examining intercepts and slopes (and their variation) that can be estimated with this type of data. Perhaps the greatest difference between HLM and analysis of variance is that analysis of variance makes use of ordinary least squares (OLS) whereas HLM makes use of maximum likelihood procedures, either full maximum likelihood (ML) or restricted maximum likelihood (REML). The default in HLM6 is restricted maximum likelihood. With least squares procedures, there are formulae that one can use to calculate the required statistics and the estimate of error is the residual. With maximum likelihood procedures, the corresponding parameters and, where applicable, their sampling error are estimated beginning with the OLS values. Maximum likelihood procedures are iterative, continually estimating the corresponding parameters (including the error terms) until the data fit the proposed model as closely as possible. It will be noted in some runs, that the fit of the model isn't always achieved, but that the fit approaches an asymptote at some value, and the decision is made that this is an acceptable estimate of fit. In

¹Preparation of this manuscript was facilitated by grant number 410-2002-0810 from the Social Science and Humanities Research Council of Canada.

any event, the values estimated with the maximum likelihood approaches differ from the OLS values. Thus, in addition to differences in the test statistics of interest, a major distinction between analysis of variance and hierarchical linear modeling is in the way in which sampling error is estimated.

The Random-Coefficients model

For this type of data, hierarchical linear modeling considers two levels of models. The model at level 1 is concerned with the regression of the dependent variable on the repeated measures factor. For the simplest case (i.e., (a), the single factor repeated measures design with only a linear component, referred to as a Linear Growth Model by Raudenbush and Bryk (2002, p 163), the level 1 equation is:

$$Y_{it} = \pi_{0i} + \pi_{1i}a_{it} + e_{it} \quad (1)$$

In this equation, the π 's refer to population values of *unstandardized* regression coefficients, where π_{0i} is the intercept and π_{1i} is the slope for individual i . The e_{it} is the error in predicting Y_{it} with this equation. In this model, π_{0i} is the true dependent variable score for individual i at $a_{it} = 0$, and π_{1i} is the rate of change over the repeated measure factor for individual i . Clearly, the rate of change for the individual will be the same regardless of how the repeated measure factor is coded, but the intercept will depend on the nature of the coding.

The level 2 equations for this model express the intercepts and slopes for the individuals as a function of other characteristics of these individuals. In the simplest case, the intercept for individual i is defined as:

$$\pi_{0i} = \beta_{00} + r_{0i} \quad (2a)$$

where:

β_{00} = mean intercept over all subjects

r_{0i} = random variation in intercepts for individuals (i.e., $\pi_{0i} - \beta_{00}$)

and the slope for the individual is defined as:

$$\pi_{1i} = \beta_{10} + r_{1i} \quad (2b)$$

where:

β_{10} = mean slope over all subjects

r_{1i} = random variation in slopes for individuals (i.e., $\pi_{1i} - \beta_{10}$)

Thus substituting equations (2a) and (2b) into equation (1) yields the full equation for the analysis:

$$Y_{it} = \beta_{00} + r_{0i} + \beta_{10}a_{1t} + r_{1i}a_{1t} + e_{it} \quad (3)$$

Note this equation indicates that the obtained Y_{it} is a function of four parameters, the mean intercept (β_{00}) and the mean slope (β_{10}) (both fixed parameters), plus two random parameters, variation of the intercepts (r_{0i}), and variation of the slopes (r_{1i}). The residual variation, (e_{it}) is what is left over after these four parameters are estimated. As we shall see later, a_{it} can be uncentered, grand mean centered, or individual mean centered depending on the nature of the model, and in turn this will influence some of the estimates and their tests of significance.

The difference between analysis of variance and hierarchical linear modeling lies in adding the level 2 equations when estimating the slopes and intercepts for each group. Using only equation (1) (which is the case if one uses multiple regression to perform an analysis of variance of these data) will yield the actual intercepts and slopes computed on the data for each individual, and define the variance due to error in terms of the residual. Using equation 3 in hierarchical linear modeling will yield maximum likelihood estimates of the mean intercept and slope over all individuals, the variation of the slopes and intercepts over the individuals, and the associated estimate of error variance. The two sets of answers will not be the same. Note that hierarchical linear modeling is concerned with estimating the unique contribution of the various parameters in the final equation. Thus, the term “hierarchical” refers to the fact that the slopes and intercepts are considered from the two points of view, level 1 and level 2. It is not hierarchical in the same sense as hierarchical multiple regression, where various predictors are entered in different steps. In hierarchical linear modeling, the equation is defined from the two points of view, but the final equation is solved in terms of the unique contribution of each parameter (i.e., SSTYPE III, or Method I (Overall & Spiegel, 1969)).

An Example from the Point of View of Analysis of Variance.

Consider the following data from a single factor repeated measures analysis of variance design.

| | T1 | T2 | T3 |
|-------|------|------|------|
| | 3 | 2 | 5 |
| | 2 | 3 | 4 |
| | 1 | 1 | 1 |
| | 3 | 3 | 2 |
| | 2 | 2 | 4 |
| | 2 | 1 | 3 |
| | 2 | 1 | 2 |
| Means | 2.14 | 1.86 | 3.00 |

With a repeated measures analysis of variance, the F-ratio for the repeated measures factor is used to assess whether the means for the factor differ more than can be attributed to chance. That is, the major data of interest are the means of each level of the repeated measures factor. If the F-ratio is significant, it suggests that the means do differ, and often the researcher

will employ post hoc tests to determine which means differ from each other. It is possible, of course, to test for linear, quadratic, etc., trends in the means, but generally the interest is simply on variability of the treatment means. The analysis of these data yields the following analysis of variance summary table.

| Source | Sum of Squares | df | Mean Square | F |
|--------------|----------------|----|-------------|--------|
| Between Ss | 12.000 | 6 | 2.000 | |
| Intercept | 114.333 | 1 | 114.333 | 57.167 |
| Within Ss | | | | |
| Time | 4.952 | 2 | 2.476 | 3.852 |
| Error (Time) | 7.714 | 12 | .643 | |

This analysis indicates that the grand mean (i.e., the Intercept) is significantly greater than 0, $F(1,6) = 57.167$, $p < .001$, and that the variation in the treatment means is marginally significant, $F(2,12) = 3.852$, $p < .051$.

The Example from the point of view of HLM.

The reason for performing an hierarchical linear modeling analysis for a repeated measure design is to determine whether there is a common form of variation across the repeated measures for all subjects as determined when the dependent measure is regressed against the repeated measure factor. There are four null hypotheses for the simplest case. One is that the mean intercept (across all Subjects) is 0; a second is that the mean slope is 0; the third is that the variance of the intercepts is 0; and the fourth is that the variance of the slopes is 0. Unless the dependent variable has a mean of 0 in the population, the first null hypothesis is of little interest. The others have psychological meaning, however. Thus, if the mean slope is found to differ significantly from 0, this indicates that there tends to be an average positive or negative relationship between the dependent variable and the repeated measures factor in the population. If the variance of the intercepts is significantly greater than 0 (note, this is a one tailed test), this would indicate that individuals differ in terms of their intercepts. Finally, if the variance of the slopes is significantly greater than 0 (again a one tailed test), this would indicate that the slope of the dependent variable against the repeated measure factor differs from individual to individual. We will restrict our example to the linear form, though the generalization to higher polynomials is straightforward (if there is a sufficient number of levels of the repeated measures factor).

The same data can be used to perform a Random Coefficients Model for a linear effect using HLM6. The data input requires two data files, one for Level 1 and one for Level 2. The Level 1 data file in SPSS would appear as follows. Note an extra variable (COV) has been added for a later analysis:

HLMREPEATEDANOVALEVEL1.sav [DataSet1] - SPSS Data Editor

| 1 : Subject | Subject | Time | DV | COV | var | var | var | var | var | var |
|-------------|---------|------|------|------|-----|-----|-----|-----|-----|-----|
| 1 | 1.00 | 1.00 | 3.00 | 7.00 | | | | | | |
| 2 | 1.00 | 2.00 | 2.00 | 5.00 | | | | | | |
| 3 | 1.00 | 3.00 | 5.00 | 9.00 | | | | | | |
| 4 | 2.00 | 1.00 | 2.00 | 6.00 | | | | | | |
| 5 | 2.00 | 2.00 | 3.00 | 8.00 | | | | | | |
| 6 | 2.00 | 3.00 | 4.00 | 8.00 | | | | | | |
| 7 | 3.00 | 1.00 | 1.00 | 5.00 | | | | | | |
| 8 | 3.00 | 2.00 | 1.00 | 4.00 | | | | | | |
| 9 | 3.00 | 3.00 | 1.00 | 5.00 | | | | | | |
| 10 | 4.00 | 1.00 | 3.00 | 6.00 | | | | | | |
| 11 | 4.00 | 2.00 | 3.00 | 6.00 | | | | | | |
| 12 | 4.00 | 3.00 | 2.00 | 5.00 | | | | | | |
| 13 | 5.00 | 1.00 | 2.00 | 5.00 | | | | | | |
| 14 | 5.00 | 2.00 | 2.00 | 6.00 | | | | | | |
| 15 | 5.00 | 3.00 | 4.00 | 7.00 | | | | | | |
| 16 | 6.00 | 1.00 | 2.00 | 5.00 | | | | | | |
| 17 | 6.00 | 2.00 | 1.00 | 3.00 | | | | | | |
| 18 | 6.00 | 3.00 | 3.00 | 6.00 | | | | | | |
| 19 | 7.00 | 1.00 | 2.00 | 4.00 | | | | | | |
| 20 | 7.00 | 2.00 | 1.00 | 3.00 | | | | | | |
| 21 | 7.00 | 3.00 | 2.00 | 4.00 | | | | | | |

The Level 2 data file would appear as follows (again with an extra variable):

HLMREPEATEDLEVEL2.sav [DataSet2] - SPSS Data Editor

| 1 : SUBJECT | SUBJECT | DV1 | SEX | var | var | var | var | var | var | var |
|-------------|---------|------|------|-----|-----|-----|-----|-----|-----|-----|
| 1 | 1.00 | 3.00 | 1.00 | | | | | | | |
| 2 | 2.00 | 2.00 | 1.00 | | | | | | | |
| 3 | 3.00 | 1.00 | 1.00 | | | | | | | |
| 4 | 4.00 | 3.00 | 2.00 | | | | | | | |
| 5 | 5.00 | 2.00 | 2.00 | | | | | | | |
| 6 | 6.00 | 2.00 | 2.00 | | | | | | | |
| 7 | 7.00 | 2.00 | 2.00 | | | | | | | |
| 8 | | | | | | | | | | |
| 9 | | | | | | | | | | |
| 10 | | | | | | | | | | |
| 11 | | | | | | | | | | |
| 12 | | | | | | | | | | |
| 13 | | | | | | | | | | |
| 14 | | | | | | | | | | |
| 15 | | | | | | | | | | |
| 16 | | | | | | | | | | |
| 17 | | | | | | | | | | |
| 18 | | | | | | | | | | |
| 19 | | | | | | | | | | |
| 20 | | | | | | | | | | |
| 21 | | | | | | | | | | |

With HLM, the emphasis is on the regression of the dependent variable against the repeated measures factor (TIME) for each individual, which in the linear case to be discussed here results in an intercept and slope for each person. In the full model, this permits tests that the mean intercept and the mean slope differ significantly from 0, and that the variances of the intercepts and slopes are each greater than 0. If the mean slope differs significantly from 0, this indicates that the repeated measures factor had a linear effect on the dependent variable. That is, the major data of interest are the intercepts and slopes for each S when the dependent variable is regressed on the repeated measures factor. If there is interest in determining whether an effect other than linear is significant, this can be assessed using orthogonal polynomials.

Running HLM

To analyze the current data, we require the following:

- (a) the dependent variable, labelled DV in the Level 1 file
- (b) the repeated measure factor, labelled TIME (and consisting of three levels).
- (c) An ID for each subject (labelled SUBJECT in both the Level 1 and Level 2 files).

Because there is no variable describing differences between individuals other than the ID, this is referred to as a Random-Coefficients Regression model. This example assumes that the data are in SPSS.SAV files. The analysis proceeds as follows:

1. Enter HLM
2. Click on **File** in the tool bar. To create a new file, click on **Make New MDM file**, and then select **Stats Package Input**.
3. This presents a window. Select **HLM2** if it is not already selected. Click on **OK**.
4. This presents another window (labelled **MAKE MDM - HLM2**). In this window:
 - a. Type in a file name with the extension **.MDM** in the pane at the upper right. This is the file that will contain the instructions for this run. If you save it as suggested later in the sequence, it will be available for future runs that might be edited, etc.
 - b. Click the drop down window for **Input File Type** and, if necessary, select **SPSS/Windows**. There is a section indicating the type of data nesting. For the type of model considered here, the **Nesting of Input Data** should have **measures within people** indicated.
 - c. Click **Browse** for Level 1.
5. This will open another window labelled **Choose Level-1 File** that presents a file source (e.g., C:\, H:\, E:\, etc, depending on your configuration). At the bottom of this window, **Files of Type** should indicate **SPSS/Windows files[*].SAV**. Click on **Open**.
6. This will return you to the window **Make MDM-HLM2**. Click on **Choose variables** (for Level-1 Specification).
7. This will produce another window labelled **Choose variables-HLM2**. This window lists all the variables in your file with two columns of boxes next to them. One is labelled **ID**, and the other **in MDM**. Put a check in the box for **ID** for the variable which indicates the subjects (SUBJECT) and checks in the boxes in the **in MDM** columns for any variable that you might wish to consider as either a predictor or an outcome variable at level 1. Click on **OK**.
8. This will return you to the window, **Make MDM-HLM2**. This window also allows you to indicate whether or not you have missing data. The default checked is **no missing data**. If you have missing data, indicate this and then click on either delete missing data when **making MDM** or on **running analyses**. Click on **Browse** for Level 2 Specification.

9. This will produce another window labelled **Choose Level-2 File** that presents a file source (e.g., C:\, H:\, E:\, etc, depending on your configuration). At the bottom of this window, **Files of Type** should indicate **SPSS Windows files [.SAV]**. Select the appropriate folder and data file. In this example, the only variable needed is SUBJECT as the ID, though HLM6 requires information for one other variable even if it is not used. Click **Open**, and this will return you to the **Make MDM-HLM2** window. Click on **Choose Variables** (for Level-2 Specification).

10. In any event, this will reintroduce the window **Choose variables-HLM2**, which lists all the variables in this file with the two columns, **ID**, and **in MDM** after each one. You are to put a check in the box for **ID** for the variable which indicates which group the subjects are in, and checks in the boxes in the **in MDM** columns for any level 2 variable that you might wish to consider as a predictor at level 2. Note there must be at least one variable checked in the MDM column, even if it is not used later, so you must put a check in a column for at least one more variable. Then click on **OK**. This returns you to the window, **Make MDM - HLM2**.

11. Click on **Save MDMT file** and save the file you just created in a folder of your choice. The computer presents a window to save files. If the default file source is not the one you want, click on the down arrow in the **Save In** window, select the source and folder of interest, click on the **File Name** pane and type in the file name. Click on **Save**. This returns you to the **Make MDM-HLM2** window.

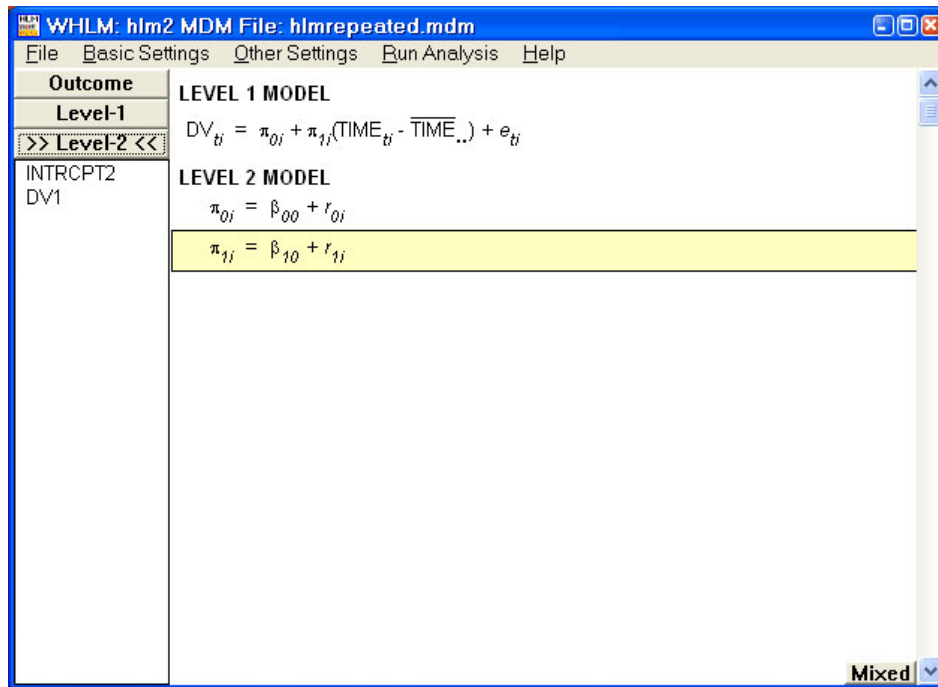
12. Click on **Make MDM**. This presents a black screen with white writing showing the descriptive statistics for all the variables chosen.

13. Click on **Done**. The computer advises you to check the statistics. Click on **OK** and then on **Done** again. This leads to the next window.

14. This presents a blank window in which you can build the model. First, click on **Level 1**, then click on the dependent variable (DV in our example) and then on **Outcome variable**. This will produce the beginning of the Level 1 equation. Next click on the repeated measure factor (TIME in our example), and then on **add variable grand mean centered**. There are actually three choices you can make here, uncentered, group mean centered, and grand mean centered (see Raudenbush & Bryk, 2002, pp. 31-35).

15. Click on Level 2, and build the Level 2 model. Our interest here is simply to study the effects of TIME, thus there are no variables to add. At this point, there are two equations at Level 2. It is possible that one of the terms, r_{0i} or r_{1i} is presented in a faded manner. If so, click on that term (to toggle it), and it will become darker. This will cause the estimate of the variance in the individual intercepts (r_{0i}) and slopes (r_{1i}) to be estimated, which you will want in order to determine whether the estimated values differ over people. If you do not want to estimate these variances, click on it to toggle it off (it becomes a lighter font).

Following is a window showing the final form of the equations.



16. The full model (labelled *Mixed Model*) is the equation containing all the parameters that are being estimated. It is as follows:

$$DV_{ti} = \beta_{00} + \beta_{10}*(TIME_{ti} - \overline{TIME_{..}}) + r_{0i} + r_{1i}*(TIME_{ti} - \overline{TIME_{..}}) + e_{ti}$$

17. HLM does not output the maximum likelihood, or the ordinary least squares, regression coefficients by default. If the maximum likelihood estimates are desired, an **SPSS.SAV** file with the estimates (and other estimates) can be obtained by clicking on **Basic Settings** in the tool bar. This will produce another window. Click on **Level 2 Residual File**. This produces another window that shows the dependent variable in a pane labelled **Possible choices**. Double click on the dependent variable and it will move it to the pane **Variables in the Residual file**. Note that at the bottom, it indicates the file (e.g., **resfil2.sav**). (In more complex models there may other variables to move. Also, if you wanted the Level 1 residual file, you could obtain this by following comparable instructions after having checked **Level 1 Residual File**). Click on **OK**.

18. To have the ordinary least squares regression coefficients printed as part of the output, Click on **Other Settings** in the tool bar and when a new window is presented turn off **Reduced Output** in that window. This will result in the OLS regression coefficients being presented in the output file along with other intermediate results. The maximum likelihood results will be the last stage presented (there are intermediate results before them).

19. To choose the type of analysis you wish to do (i.e., REML or ML), click on **Other Settings** in the tool bar, and then on **Estimation Settings**. The default setting is Restricted Maximum Likelihood (REML), which we will use in this example, but you could choose Maximum Likelihood (ML) if you prefer. The answers differ as well as some of the output depending on which one you select.

20. When the model is ready, click on **Run Analysis**. If you have not saved the run, the computer will remind you of this. If you choose **Save and Run**, it will produce another window, **Save Command File**. Type in the file name and it will save it.

21. To view the output, click **File** and **View Output**.

The Output

The HLM analysis of the data yields the following output using REML. If the alterations described in steps 17 and 18 are followed, HLM6 will yield both the OLS and REML estimates of the intercept and slope for each individual. For this example, that information is as follows:

| Ordinary Least Squares Estimates | | | Maximum Likelihood Estimates | |
|----------------------------------|----------|------------|------------------------------|------------|
| Level-2 Unit | INTRCPT1 | TIME slope | INTERCEPT | TIME slope |
| 1 | 3.33333 | 1.00000 | 3.063 | .819 |
| 2 | 3.00000 | 1.00000 | 2.861 | .711 |
| 3 | 1.00000 | 0.00000 | 1.432 | -.058 |
| 4 | 2.66667 | -0.50000 | 2.335 | .428 |
| 5 | 2.66667 | 1.00000 | 2.659 | .603 |
| 6 | 2.00000 | 0.50000 | 2.147 | .329 |
| 7 | 1.66667 | 0.00000 | 1.837 | .163 |
| Means | 2.33333 | .42857 | 2.333 | .42857 |

It should be noted that the OLS values are simply those that you would obtain if you were to perform a regression analysis on each individual separately. In HLM6, these values are used as initial values in the iteration process. Inspection of the OLS and REML estimates for each of the 7 individuals will reveal that there are differences in the corresponding individual values in the two sets though the means are identical. Not shown here are the tests of significance obtained from the OLS analysis, though they are available in the output if the “Reduced Output” option is turned off as described in Step 18 (though, researchers do not generally output this information).

Following is the output for this analysis, with explanatory comments (indicated by ***) interspersed.

SPECIFICATIONS FOR THIS HLM2 RUN

Problem Title: no title

The data source for this run = C:\PSYCH540\HLMREPEATED.MDM

The command file for this run = whlmtemp.hlm

Output file name = C:\PSYCH540\hlm2.txt

The maximum number of level-1 units = 21

The maximum number of level-2 units = 7

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 DV

The model specified for the fixed effects was:

```

-----
Level-1          Level-2
Coefficients      Predictors
-----
    INTRCPT1, P0  INTRCPT2, B00
%  TIME slope, P1  INTRCPT2, B10

```

'%' - This level-1 predictor has been centered around its grand mean.

The model specified for the covariance components was:

Sigma squared (constant across level-2 units)

Tau dimensions

INTRCPT1

TIME slope

Summary of the model specified (in equation format)

Level-1 Model

$$Y = P0 + P1*(TIME) + E$$

Level-2 Model

$$P0 = B00 + R0$$

$$P1 = B10 + R1$$

*** The following two lines indicate that it took 3190 iterations to produce this solution. A minimum value of the fit function was not obtained, but rather the function was deemed to have reached an asymptote. Some programs issue a warning that the solution may not be valid, but such warnings are not issued in HLM6.***

Iterations stopped due to small change in likelihood function

***** ITERATION 3190 *****

*** The following indicates that the variance due to error (E) in the equation was .63937.***

Sigma_squared = 0.63937

*** The tau matrix is a covariance matrix for the individual intercepts and slopes. The diagonal values are the estimates of the population variances of the intercepts and slopes respectively, while the value in the off-diagonal is the estimate of the covariance of the intercepts and slopes. This matrix is also presented in standard score form and hence is the correlation matrix obtained from the covariance matrix by dividing by the product of the associated standard deviations. Following are the two matrices:***

Tau

| | | |
|-------------|---------|---------|
| INTRCPT1,P0 | 0.46416 | 0.24800 |
| TIME,P1 | 0.24800 | 0.13292 |

Tau (as correlations)

| | | |
|-------------|-------|-------|
| INTRCPT1,P0 | 1.000 | 0.998 |
| TIME,P1 | 0.998 | 1.000 |

*** This section reports the reliability estimates for the Level 1 intercepts and slopes. Reliability is the ratio of “true” parameter variance to the total observed variance, calculated separately for each individual and averaged over all individuals. High reliabilities indicate a high level of precision in the estimates. In this case, the reliabilities are low (particularly for the slopes), probably because the sample size as well as the number of levels of the repeated measures factor is so small. With low precision, significant results are less likely. ***

Random level-1 coefficient Reliability estimate

| | |
|--------------|-------|
| INTRCPT1, P0 | 0.685 |
|--------------|-------|

Hence, each of these is a one-tailed test. Note that the variance components are in fact the diagonal values in the tau matrix presented above. The tests of significance in this case make use of a Chi-square statistic with degrees of freedom equal to the number of individuals minus 1.***

Final estimation of variance components:

| Random Effect | Standard Deviation | Variance Component | df | Chi-square | P-value |
|----------------|--------------------|--------------------|----|------------|---------|
| INTRCPT1, R0 | 0.68129 | 0.46416 | 6 | 18.76853 | 0.005 |
| TIME slope, R1 | 0.36458 | 0.13292 | 6 | 6.92648 | 0.327 |
| level-1, E | 0.79960 | 0.63937 | | | |

Statistics for current covariance components model

Deviance = 56.944799

Number of estimated parameters = 4

Interpretation

Interpretation of these data would focus on the two primary tables. The table of Final estimation of fixed effects indicates that the mean intercept is 2.3333. Because the Repeated measures factor was grand mean centered and because there are no missing observations, this value corresponds to the grand mean of the data. The test of significance indicates that this mean differs significantly from 0, $t(6) = 7.501$, not particularly surprising since all the dependent values were greater than 0. Note, if the repeated measures factor had not been centered, the “intercept” would differ from that obtained here, as would the test of significance. The table also presents the coefficient for the slope, as .428571, which because there are no missing values corresponds to the mean of the REML estimates of the individual slopes shown earlier. This value does not differ significantly from 0, $t(6) = 1.685$. If it were significant, we could conclude that on average, the dependent variable increases by .428571 for each increment in the Repeated Measures factor.

The table of Final estimation of variance components presents three estimates. The first is the variance of the intercepts across the individuals estimated at .46416 (note this is the value in the diagonal of the Tau matrix), and tested for significance using a chi-square statistic, $\chi^2(6) = 18.76853$, $p < .005$. This indicates that the individual intercepts differ more than would be expected on the basis of chance, and since these intercepts correspond to the individual means, it suggests that some individuals obtain higher scores on the dependent variable than others. The second estimate is for the variance of the individual slopes (.13292) and is the value in the diagonal of the Tau matrix. The test is not significant, $\chi^2(6) = 6.92648$, indicating that there is no evidence of individual variation in the slopes. The final statistic estimated is the residual variance shown to be .63937, referred to as sigma squared on page 11. This value is a measure of the variance not accounted for by the analysis, and is not tested for significance in HLM.

Other Possible Analyses of These Data.

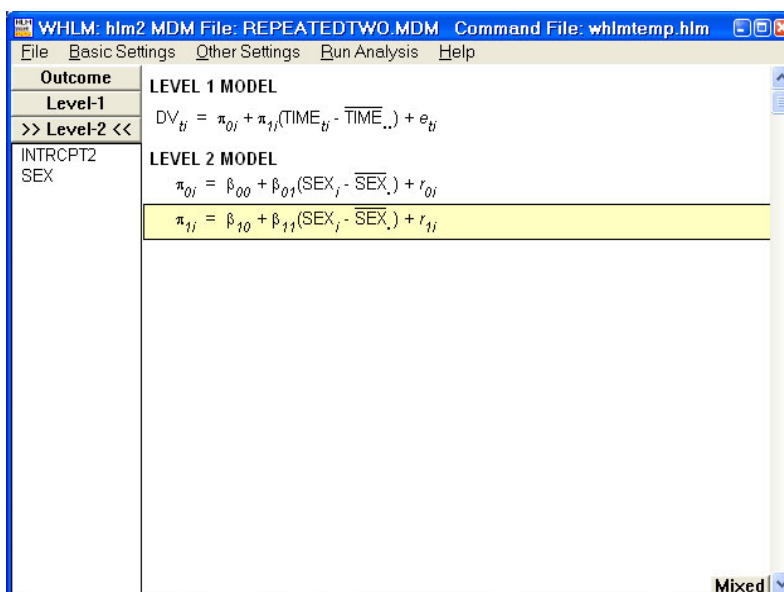
There are other HLM analyses that could be done with these data. For example, one could omit the TIME variable in the Level 1 equation. This would result in the program not estimating slopes for each subject, and thus the intercept would correspond to the individual's mean. Obviously, there would be no test of the mean and variance of the slope, and the tests of the intercept would refer to the grand mean and variance of the means as estimated by maximum likelihood. Another example, might include the intercept and slope as in the present example, but if either r_{0i} or r_{1i} or both were not estimated, this is equivalent to setting them equal to 0, and thus corresponding tests of the variances would not be calculated. Of course, some of the values would be changed with these variants, because the model has been changed, so the researcher should be aware of the precise nature of the model presented and explain it in any subsequent report.

Generalizing the Analyses

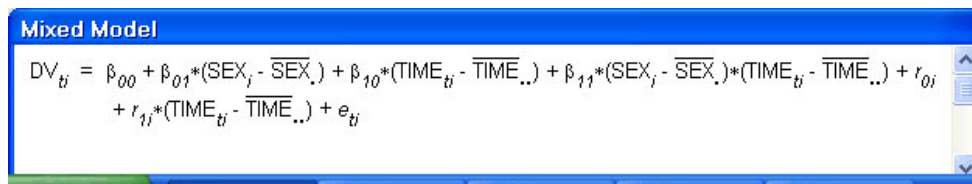
There are a number of different analyses that can be done using HLM. In this section, we briefly consider two more, using the sample data, and limiting us to the general two-level model.

The equivalent of a Split Plot Factorial Design.

The first analysis to be considered is the equivalent of a split plot factorial design with Sex as the Between Subjects factor and Time as the Within Subjects factor. The following shows the HLM6 setup:



This produces the following mixed model.



$$DV_{ij} = \beta_{00} + \beta_{01}*(SEX_{ij} - \overline{SEX}_{..}) + \beta_{10}*(TIME_{ij} - \overline{TIME}_{..}) + \beta_{11}*(SEX_{ij} - \overline{SEX}_{..})*(TIME_{ij} - \overline{TIME}_{..}) + r_{0i} + r_{1j}*(TIME_{ij} - \overline{TIME}_{..}) + e_{ij}$$

The output for the fixed and random effects are presented in the following two tables.

Final estimation of fixed effects:

| Fixed Effect | Coefficient | Standard Error | T-ratio | Approx. d.f | P-value |
|--------------------|-------------|----------------|---------|-------------|---------|
| For INTRCPT1, P0 | | | | | |
| INTRCPT2, B00 | 2.333333 | 0.337216 | 6.919 | 5 | 0.000 |
| SEX, B01 | -0.194444 | 0.681422 | -0.285 | 5 | 0.787 |
| For TIME slope, P1 | | | | | |
| INTRCPT2, B10 | 0.428571 | 0.258889 | 1.655 | 5 | 0.158 |
| SEX, B11 | -0.416667 | 0.523144 | -0.796 | 5 | 0.462 |

This table provides information about four estimates. First, the mean intercept is 2.3333 and it differs significantly from 0, $t(5) = 6.919$, $p < .0004$. Note that this is the grand mean of the DV and it is identical to that obtained in the first example. Note too that the degrees of freedom are one less than they were for that example. Second, note that the mean slope is the same as for the first example and that the degrees of freedom are also one less. This value does not differ significantly from 0, $t(5) = 1.655$, $p < .158$, but if it did it would indicate that the rate of change of the DV over TIME was .428571. Third, the regression of the intercepts on Sex is -.194444, but it is not significant $t(5) = -.285$, *ns*. If it were, it would indicate that there is a negative linear relationship between the intercepts and Sex. In this case, it would indicate that those coded 1 had higher intercepts than those coded 2 (i.e., the two sexes differed on the intercepts). Finally, the regression of the slopes on Sex was also negative, -.416667, and not significant, $t(5) = -.796$, *ns*, but if it had been it would indicate a negative relationship between the slopes and Sex.

Final estimation of variance components:

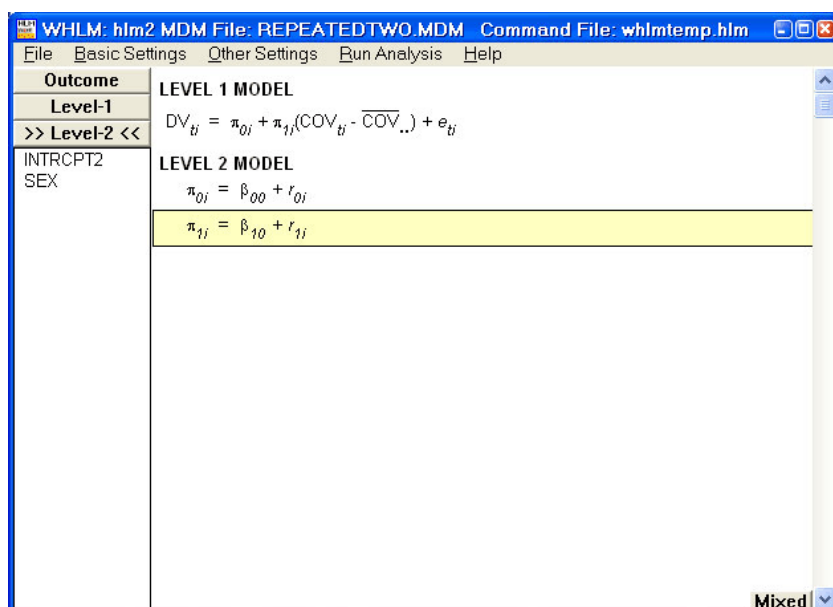
| Random Effect | Standard Deviation | Variance Component | df | Chi-square | P-value |
|----------------|--------------------|--------------------|----|------------|---------|
| INTRCPT1, R0 | 0.75718 | 0.57331 | 5 | 17.67123 | 0.004 |
| TIME slope, R1 | 0.36760 | 0.13513 | 5 | 5.73795 | 0.332 |
| level-1, E | 0.81735 | 0.66807 | | | |

This table tests the variance of the intercepts and slopes over the subjects. The results

indicate that the variance of the intercepts, .57331, is significantly greater than 0, $\chi^2(5) = 17.67123$, $p < .004$ suggesting that there are true individual differences in the intercepts. The variance in the slopes, .13513, is not significantly greater than 0, $\chi^2(5) = 5.73795$, suggesting that the individuals have similar slopes. Note, furthermore, that although the mean intercept and slope were the same as for the first example, these variance estimates are not, and that the degrees of freedom were reduced by 1 because of the additional variable, Sex.

An Analysis with a Covariate Varying Over Time.

This analysis also makes use of repeated measures over time, but in this case the interest is in whether another Level 1 variable that covaries with Time has an effect on the dependent variable. Following is the setup for HLM6.



This yields the following Mixed model.

The Mixed Model window displays the following equation:

$$DV_{ti} = \beta_{00} + \beta_{10} * (\text{COV}_{ti} - \overline{\text{COV}}_{..}) + r_{0i} + r_{1i} * (\text{COV}_{ti} - \overline{\text{COV}}_{..}) + e_{ti}$$

The output for the two major tables is as follows:

Final estimation of fixed effects:

| Fixed Effect | Coefficient | Standard Error | T-ratio | d.f. | Approx. P-value |
|--------------|-------------|----------------|---------|------|-----------------|
|--------------|-------------|----------------|---------|------|-----------------|

| | | | | | | |
|-----|---------------|----------|----------|--------|---|-------|
| For | INTRCPT1, P0 | | | | | |
| | INTRCPT2, B00 | 2.330653 | 0.182884 | 12.744 | 6 | 0.000 |
| For | COV slope, P1 | | | | | |
| | INTRCPT2, B10 | 0.683154 | 0.077186 | 8.851 | 6 | 0.000 |

This table presents tests of two fixed effects, the mean intercept and the mean slope. Each are significant, $t(6) = 12.744$ and 8.851 , respectively. Note that, as for the first example, the degrees of freedom are 6, but now the mean intercept and slope are different. This is to be expected, of course, because the covariate at each time period is the independent variable rather than the time variable itself.

Final estimation of variance components:

| Random Effect | Standard Deviation | Variance Component | df | Chi-square | P-value |
|---------------|--------------------|--------------------|----|------------|---------|
| INTRCPT1, R0 | 0.42537 | 0.18094 | 6 | 12.57221 | 0.050 |
| COV slope, R1 | 0.00772 | 0.00006 | 6 | 4.39839 | >.500 |
| level-1, E | 0.39922 | 0.15938 | | | |

This table shows that the variance of the intercepts over individuals is .18094, which is significantly greater than 0, $\chi^2(6) = 12.57221$, $p < .050$, suggesting that there are true individual differences in performance. It also shows that the variance of the slopes, .00006, is not significantly greater than 0, $\chi^2(6) = 4.39839$ suggesting that the rate of change in the dependent variable as a function of the covariate is comparable for the individuals. Again, note that the degrees of freedom are 6, as in the first example.

There are, of course, many other possible analyses that can be conducted in HLM. In all cases, the estimates of interest are the intercepts and slopes. Seldom are the intercepts and slopes for the groups reported; the fixed effects (i.e., the mean values) reflect the general effect. Where a random effect is significant, one might wish to display the group values in a figure.

References

- Gardner, R. C. (2007). *Hierarchical Linear Modeling: A Primer (Persons within Groups)*. Unpublished manuscript, University of Western Ontario, London, Canada..
- Overall, J.E., & Spiegel, D.K. (1969). Concerning least squares analysis of experimental data. *Psychological Bulletin*, 72, 311-322.
- Raudenbush, S.W. and Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage