2 Working with HLM2

Data analysis by means of the HLM2 program will typically involve three stages:

1. construction of the "MDM file" (the multivariate data matrix);
2. execution of analyses based on the MDM file; and
3. evaluation of fitted models based on a residual file.

We describe each stage below and then illustrate a number of special options. Data collected from a High School & Beyond (HS&B) survey on 7,185 students nested within 160 US high schools, as described in Chapter 4 of *Hierarchical Linear Models*, will be used for demonstrations.

2.1 Constructing the MDM file from raw data

We assume that a user has employed a standard computing package to clean the data, make necessary transformations, and conduct relevant exploratory and descriptive analyses. We also recommend exploratory graphical analyses within HLM prior to model building as described in detail in Section 12.1 of this manual.

The first task in using HLM2 is to construct the Multivariate Data Matrix (MDM) from raw data or from a statistical package. We generally work with two raw data files: a level-1 file and a level-2 file. Both files must be sorted by the level-2 ID (It is possible, however, to build the MDM file from the level-1 file above, though this option is not suggested when the level-1 file is very large. The level-1 file must be sorted by level-2 ID. The level-1 file name will be selected as both the level-1 and level-2 file).

For the HS&B example, the level-1 units are students and the level-2 units are schools. The two files are linked by a common level-2 unit ID, school id in our example, which must appear on every level-1 record. In constructing the MDM file, the HLM program will compute summary statistics based on the level-1 unit data and store these statistics together with level-2 data.

The procedure to create a MDM file consists of three major steps. The user needs to

- Inform HLM of the input and MDM file type.
- Supply HLM with the appropriate information for the data, the command and the MDM files.
- Check if the data have been properly read into HLM.
2.2 Executing analyses based on the MDM file

Once the MDM file is constructed, all subsequent analyses will be computed using the MDM file as input. It will therefore be unnecessary to read the larger student-level data file in computing these analyses. The efficient summary of data in the MDM file leads to faster computation. The MDM file is like a "system file" in a standard computing package in that it contains not only the summarized data but also the names of all of the variables.

Model specification has three steps:

- Specifying the level-1 model, which defines a set of level-1 coefficients to be computed for each level-2 unit.
- Specifying a level-2 structural model to predict each of the level-1 coefficients.
- Specifying the level-1 coefficients to be viewed as random or non-random.

The output produced from these analyses includes:

- Ordinary least squares and generalized least squares results for the fixed coefficients defined in the level-2 model.
- Estimates of variance and covariance components and approximate chi-square tests for the variance components.
- A variety of auxiliary diagnostic statistics.

Additional output options and hypothesis-testing procedures may be selected.

2.3 Model checking based on the residual file

After fitting a hierarchical model, it is wise to check the tenability of the assumptions underlying the model:

- Are the distributional assumptions realistic?
- Are results likely to be affected by outliers or influential observations?
- Have important variables been omitted or non-linear relationships been ignored?

These questions and others can be addressed by means of analyses of the HLM residual files. A level-1 residual file includes:

- The level-1 residuals (discrepancies between the observed and fitted values).
- Fitted values for each level-1 unit (that is, values predicted on the basis of the model).
- The observed values of all predictors included in the model.
- Selected level-2 predictors useful in exploring possible relationships between such predictors and level-1 residuals.
A level-2 residual file includes:

- Fitted values for each level-1 coefficient (that is, values predicted on the basis of the level-2 model).
- Ordinary least squares (OL) and empirical Bayes (EB) estimates of level-2 residuals (discrepancies between level-1 coefficients and fitted values).
- Empirical Bayes coefficients, which are the sum of the EB estimates and the fitted values.
- Dispersion estimates useful in exploring sources of variance heterogeneity at level 1.
- Expected and observed Mahalanobis distance measures useful in assessing the multivariate normality assumption for the level-2 residuals.
- Selected level-2 predictors useful in exploring possible relationships between such predictors and level-2 residuals.
- Posterior variances.

For HLM2 FML analyses, there is an additional set of posterior variances. See Chapter 9 in *Hierarchical Linear Models* for a full discussion of these methods.

### 2.4 Windows, interactive, and batch execution

Formulation and testing of models using HLM programs can be achieved via Windows, interactive, or batch modes. Most PC users will find the Windows mode preferable. This draws on the visual features of Windows while preserving the speed of use associated with a command-oriented (batch) program. Non-PC users have the choice of interactive and batch modes only. Interactive execution guides the user through the steps of the analysis by posing questions and providing a menu of options. In this chapter, we employ the Windows mode for all the examples. Descriptions and examples on how to use HLM2 in interactive and batch modes are given in Appendix A.

### 2.5 An example using HLM2 in Window mode

Chapter 4 in *Hierarchical Linear Models* presents a series of analyses of data from the HS&B survey. A level-1 model specifies the relationship between student socioeconomic status (SES) and mathematics achievement in each of 160 schools; at level-2, each school’s intercept and slope are predicted by school sector (Catholic versus public) and school mean social class. We reproduce one analysis here (see Table 4.5 in *Hierarchical Linear Models*, p. 82).

#### 2.5.1 Constructing the MDM file from raw data

PC users may construct the MDM file directly from different types of input files including SPSS, ASCII, SAS, SYSTAT, and STATA, or indirectly from many additional types of data file formats through the third-party software module included in the HLM program.
Non-PC users may construct the MDM file with one of the following types of input files: ASCII data files, SYSTAT data files, or SAS V5 transport files.

In order for the program(s) to correctly read the data, the IDs need to conform to the following rules:

1. For ASCII data the ID variables must be read in as character (alphanumeric). These IDs are indicated by the A field(s) in the format statement. For all other types of data, the ID may be character or numeric.
2. The level-1 cases must be grouped together by their respective level-2 unit ID. To assure this, sort the level-1 file by the level-2 ID field prior to entering the data into HLM2.
3. If the ID is numeric, it must be in the range \((-10^{12} + 1)\) to \(+10^{12} + 1\) \(\text{(i.e. 12 digits).} \) Although the ID may be a floating point number, only the integer part is used.
4. If the ID variable is character, the length must not exceed 12 characters. Furthermore, the IDs at a given level must all be the same length. \text{This is often a cause of problems.} \text{For example, imagine your data has IDs ranging from "1" to "100". You will need to recreate the IDs as "001" to "100". In other words, all spaces (blank characters) should be coded as zeros.}
5. For non-ASCII files, the program can only properly deal with numeric variables (with the exception of character ID variables). Other data types, such as a "Date format", will not be processed properly.
6. For non-ASCII files with missing data, one should only use the "standard" missing value code. Some statistical packages (SAS, for example) allow for a number of missing value codes. The HLM modules are incapable of understanding these correctly, thus these additional missing codes need to be recoded to the more common "." (period) code.

\subsection{SPSS file input}

We first illustrate the use of SPSS file input and then consider input from ASCII data files. Data input requires a level-1 file and a level-2 file.

\textbf{Level-1 file.} For our HS&B example data, the level-1 file (HSB1.SAV) has 7,185 cases and four variables (not including the SCHOOL ID). The variables are:

- MINORITY, an indicator for student ethnicity \((1 = \text{minority}, 0 = \text{other})\)
- FEMALE, an indicator for student gender \((1 = \text{female}, 0 = \text{male})\)
- SES, a standardized scale constructed from variables measuring parental education, occupation, and income
- MATHACH, a measure of mathematics achievement

Data for the first ten cases in HSB1.SAV are shown in Fig. 2.1.

\text{Note: level-1 cases must be grouped together by their respective level-2 unit ID. To assure this, sort the level-1 file by the level-2 unit ID field prior to entering the data into HLM2.
Figure 2.1 First ten cases in HSB1.SAV

Level-2 file. At level 2, the illustrative data set HSB2.SAV consists of 160 schools with 6 variables per school. The variables are:

- SIZE (school enrollment)
- SECTOR (1 = Catholic, 0 = public)
- PRACAD (proportion of students in the academic track)
- DISCLIM (a scale measuring disciplinary climate)
- HIMNTY (1 = more than 40% minority enrollment, 0 = less than 40%)
- MEANSES (mean of the SES values for the students in this school who are included in the level-1 file)

The data for the first ten schools are displayed in Fig 2.2.

Figure 2.2 First ten cases in HSB2.SAV

As mentioned earlier, the construction of an MDM file consists of three major steps. This will now illustrated with the HS&B example.
To inform HLM of the input and MDM file type

1. At the WHLM window, open the File menu.
2. Choose Make new MDM file...Stat package input (see Figure 2.3). A Select MDM type dialog box opens (see Figure 2.4).
3. Select HLM2 and click OK. A Make MDM - HLM2 dialog box will open (see Figure 2.5).

To supply HLM with appropriate information for the data, the command, and the MDM files:

1. Select SPSS/Windows from the Input File Type pull-down menu (see Figure 2.5).
2. Click Browse in the Level-1 specification section to open an Open Data File dialog box.
3. Open a level-1 SPSS system file in the HLM folder (HSB1.SAV in our example). The Choose Variables button will be activated.
4. Click Choose Variables to open the Choose Variables - HLM2 dialog box and choose the ID and variables by clicking the appropriate check boxes (See Figure 2.6). To deselect, click the box again.
5. Select the options for missing data in the level-1 file (there is no missing data in HSB1.SAV; see Section 2.6 for details).
6. Click the selection button for measures within persons for the type of nesting of input data if the level-1 data consist of repeated measures or item responses. With this selection, WHLM will use in its displays and output model notations that match those used in Hierarchical Linear Models for studies on individual change and latent variables (Chapters 6 and 11). The default type is persons within groups. It is generally used when the level-1 data are comprised of cross-sectional measures. With this option, WHLM will use model notations that correspond to those used for applications in organization research (Chapters 4 and 5).
7. Click Browse in the Level-2 specification section to open an Open Data File dialog box.
8. Open a level-2 SPSS system file in the HLM folder (HSB2.SAV in our example). The Choose Variables button below Browse will be activated.
9. Click Choose Variables to open the Choose Variables - HLM2 dialog box and choose the ID and variables by clicking the appropriate check boxes (see Figure 2.7).
10. Enter a name for the MDM file in the MDM file name box (for example, HSB.MDM).
11. Click Save mdmt file in the MDM template file section to open a Save MDM template file dialog box. Enter a name for the MDM file (for example, HSBSPSS.MDM). Click Save to save the file. The command file saves all the input information entered by the user. It can be re-opened by clicking the Open mdmt file button (see Figure 2.5). To make changes to an existing MDM file, click the Edit mdmt file button.
12. Note that HLM will also save the input information into another file called CREATMDM.MDM when the MDM is created.
13. Click the Make MDM button. A screen displaying the prompts and responses for MDM creation will appear.
Figure 2.3  WHLM window

Figure 2.4  Select MDM type dialog box
Figure 2.5 Make MDM - HLM2 dialog box

Figure 2.6 Choose Variables - HLM2 dialog box for the level-1 file, HSB1.SAV
Figure 2.7  Choose variables - HLM2 dialog box for the level-2 file, HSB2.SAV

Figure 2.8  Descriptive Statistics for the MDM file, HSB.MDM
To check whether the data have been properly read into HLM

1. Click Check Stats to display and check the level-1 and level-2 descriptive statistics (See Figure 2.8). Pay particular attention to the N column. It is not an uncommon mistake to forget to sort by the ID variable, which can lead to a lot (or most) of the data not being processed. Close the Notepad window when done. Use the Save As option to give it a new name if later use of this file is anticipated.

2. Click Done. The WHLM window displays the type and name on its title bar (hlm2 & HSB.MDM) and the level-1 variables on a drop-down menu (See Figure 2.9).

![Figure 2.9 WHLM: hlm2 MDM File window for HSB.MDM](image)

2.5.1.2 ASCII file input

Below is the procedure for creating a multivariate data matrix file with input from ASCII files.

To inform HLM of the input and MDM file type

1. At the WHLM window, open the File menu.
2. Choose Make new MDM file...ASCII input. A Select MDM type dialog box opens.
3. Select HLM2 (see Figure 2.4) and click OK. A Make MDM File – HLM2 will open (see Figure 2.10).

To supply HLM with appropriate information for the data, the command, and the MDM files

1. Click Browse in the Level-1 specification section to open an Open Data File dialog box. Open a level-1 ASCII data file in the HLM examples folder (HSB1.DAT in our example). The file name (HSB1.DAT) appears in the Level-1 File Name box.
2. Enter the number of variables into the Number of Variables box (4 in our example) and the data entry format in the Data Format box (A4.4F12.3 in our example).
Note that the ID is included in the format statement, but excluded in the Number of Variables box. Rules for input format statements are given in Section A.2 in Appendix A.

Figure 2.10  Make MDM – HLM2 dialog box

3. Click Labels to open the Enter Variable Labels dialog box.
4. Enter the variable names into the boxes (MINORITY, FEMALE, SES, MATHACH for our example, see Figure 2.11). Click OK.
5. Click the Missing Data button to enter level-1 missing data info (there is no missing data in HSB1.DAT; see Section 2.6 for details).
6. Click Browse in the Level-2 specification section to open an Open Data File dialog box. Open a level-2 ASCII data file in the HLM folder (HSB2.DAT in our example). The file name (HSB2.DAT in our example) will appear in the in the Level-2 File Name box.
7. Enter the number of variables into the Number of Variables box (6 in our example) and the data entry format in the Data Format box (A4.6F12.3 in our example).
8. Click Labels to open the Enter Variable Labels dialog box for the level-2 variables.
9. Enter the variable names into the Variable boxes (SIZE, SECTOR, PRACAD, DISCLIM, HIMINTY, MEANSES in our example, see Figure 2.12). Click OK.
10. Enter an MDM file name in the MDM File Name box (for example, HSB.MDM).
11. Click Save mdmf file in the MDM template file section to open a Save MDM template file dialog box. Enter a name for the MDT file (for example, HSBASCIL.MDT). Click Save to save the file. The command file saves all the input information entered by the user. It can be re-opened or changed by clicking either the Open mdmf file or the Edit mdmf file button (see Figure 2.10).
\textbf{Figure 2.11} Enter Variable Labels dialog box for level-1 file, HSB1.DAT

\textbf{Figure 2.12} Enter Variable Labels dialog box for level-2 file, HSB2.DAT

To check whether the data have been properly read into HLM

The procedure is the same as for SPSS file input (see Section 2.5.1.1 for a complete description).
2.5.1.3 SAS transport, SYSTAT, STATA file input and other formats for raw data

For SAS transport, SYSTAT or STATA file input, a user selects either SAS 5 transport, SYSTAT or STATA from the Input File Type drop-down menu as appropriate to open the Open Data File dialog box. With the third-party software module included in the current version, HLM will read data from EXCEL, LOTUS and many other formats. Select Anything else from the Input File Type drop-down menu before clicking on the Browse button in the input file specifications sections. If the data type is set on the File, Preferences screen, the program will default to your selected type for both input data and residual files.

2.5.2 Executing analyses based on the MDM file

Once the MDM file is constructed, it can be used as input for the analysis. As mentioned earlier, model specification has three steps:

- Specification of the level-1 model. In our example, we shall model mathematics achievement (MATHACH) as the outcome, to be predicted by student SES. Hence, the level-1 model will have two coefficients: the intercept and the SES-MATHACH slope.
- Specification of the level-2 prediction model. We shall predict each school’s intercept by school SECTOR and MEANSES in our example. Similarly, SECTOR and MEANSES will predict each school’s SES-MATHACH slope.
- Specification of level-1 coefficients as random or non-random. We shall model both the intercept and the slope as having randomly varying residuals. That is, we are assuming that the intercept and slope vary not only as a function of the two predictors, SECTOR and MEANSES, but also as a function of a unique school effect. The two school residuals (e.g., for the intercept and slope) are assumed sampled from a bivariate normal distribution.

The procedure for executing analyses based on the MDM file is described below.

**Step 1: To specify the level-1 prediction model**

1. From the HLM window, open the File menu.
2. Choose Create a new model using an existing MDM file to open an Open MDM File dialog box. Open an existing MDM file (HSB.MDM in our example). The name of the MDM file will be displayed on the title bar of the main window. A listbox for level-1 variables (>>Level-1<<) will appear (see Figure 2.13).
3. Click on the name of the outcome variable (MATHACH in our example). Click Outcome variable (see Figure 2.13). The specified model will appear in equation format.