



Exploratory analysis of potential level-2 predictors using the HSB data

1. Description of the data	1
2. Creating the command file	3
3. Interpreting the output.....	5

1. Description of the data

High School and Beyond (HS&B) is a national longitudinal study originally funded by the United States Department of Education’s National Center for Education Statistics (NCES) as a part of their longitudinal studies program. Its purpose was to document the “educational, vocational, and personal development of young people following them over time as they begin to take on adult roles and responsibilities”. Here a subset of the data representing 160 schools and a total of 7,185 students is used to illustrate the fitting and interpretation of a two-level hierarchical linear model.

This example is the fifth in a set of seven examples based on these data described on the HLM Support page and used the MDM file created in the first example (**HSB.MDM**).

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 = minority, 0 = other)
- FEMALE, an indicator for student gender (1 = female, 0 = 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. 1.1.

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.

	id	minority	female	ses	mathach
1	1224	0	1	-1.528	5.876
2	1224	0	1	-.588	19.708
3	1224	0	0	-.528	20.349
4	1224	0	0	-.668	8.781
5	1224	0	0	-.158	17.898
6	1224	0	0	.022	4.583
7	1224	0	1	-.618	-2.832
8	1224	0	0	-.998	.523
9	1224	0	1	-.888	1.527
10	1224	0	0	-.458	21.521

Figure 1.1 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)

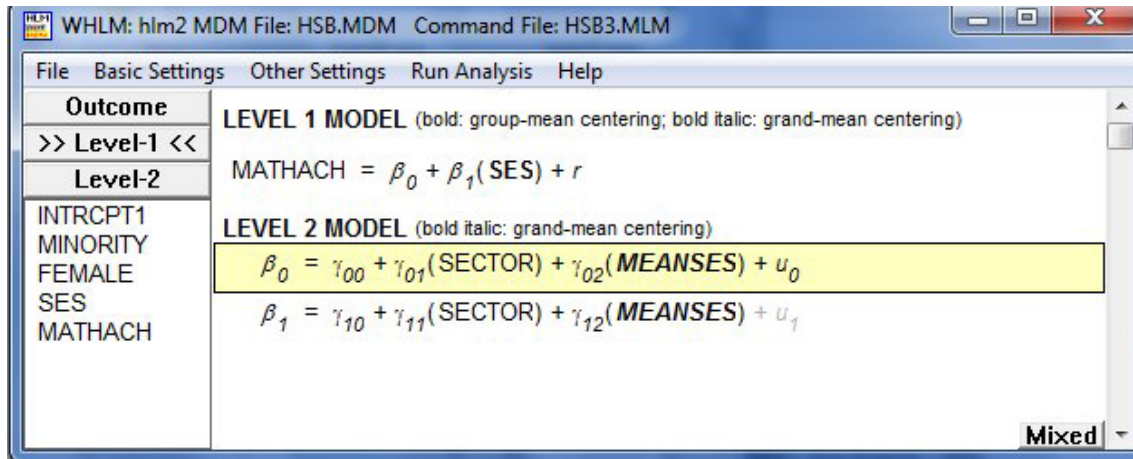
The data for the first ten schools are displayed in Fig 1.2.

	id	size	sector	pracad	disclim	himinty	meanses
1	1224	842	0	.350	1.597	0	-.428
2	1288	1855	0	.270	.174	0	.128
3	1296	1719	0	.320	-.137	1	-.420
4	1308	716	1	.960	-.622	0	.534
5	1317	455	1	.950	-1.694	1	.351
6	1358	1430	0	.250	1.535	0	-.014
7	1374	2400	0	.500	2.016	0	-.007
8	1433	899	1	.960	-.321	0	.718
9	1436	185	1	1.000	-1.141	0	.569
10	1461	1672	0	.780	2.096	0	.683

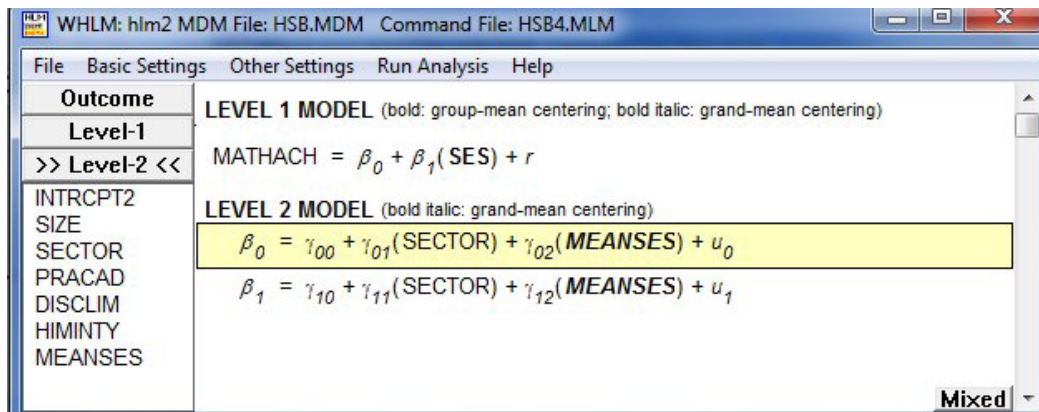
Figure 1.2 First ten cases in HSB2.SAV

2. Creating the command file

Users may wish to compute “t-to-enter” statistics for potential additional level-2 predictors to guide specification of subsequent models. The command file created in the previous example is used as point of departure.



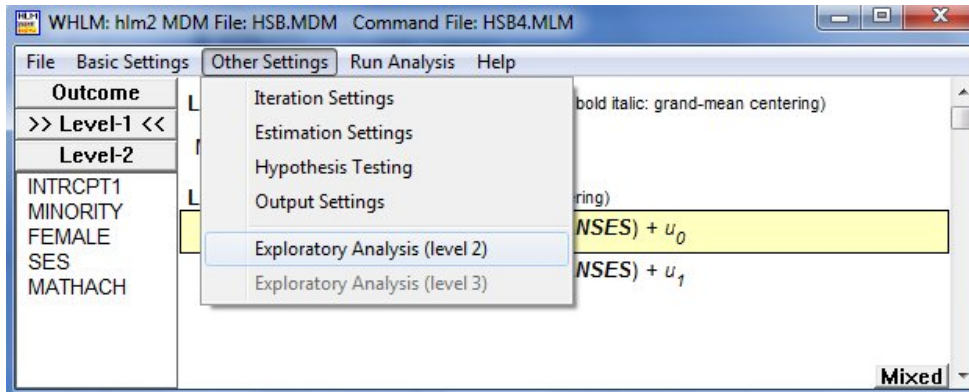
As a first step, activate the random effect on the slope equation (u_1) by double-clicking on it.



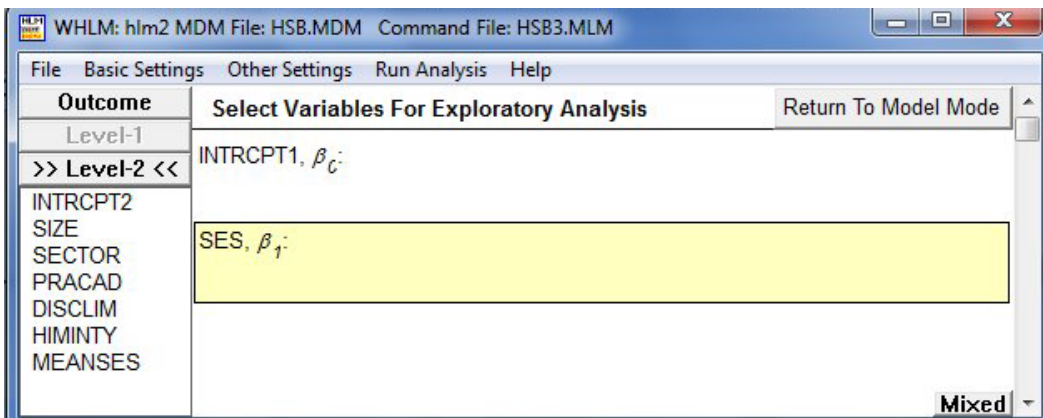
In addition to the predictor SECTOR and MEANSES already used in the level-2 model, three more school level predictors are available in the MDM file:

- PRACAD (proportion of students in the academic track)
- DISCLIM (a scale measuring disciplinary climate)
- HIMINTY (1 = more than 40% minority enrollment, 0 = less than 40%)

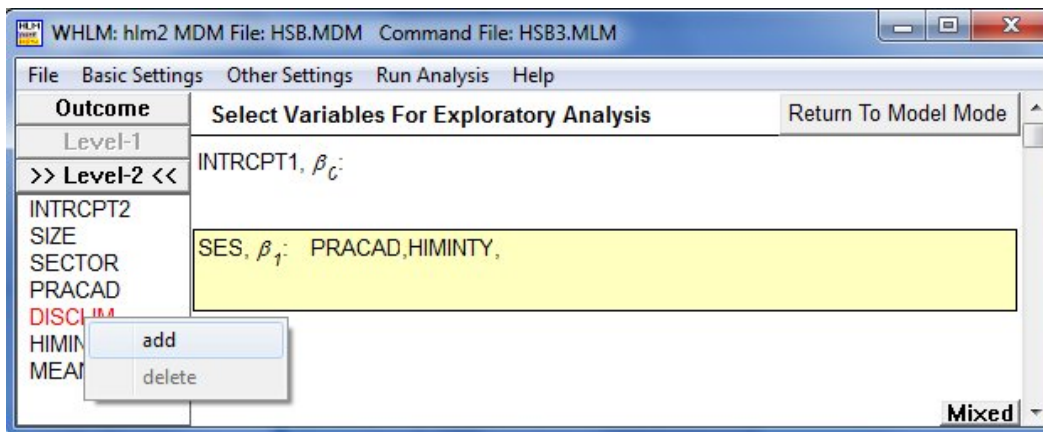
To obtain “t-to-enter” statistics for these predictors, select the **Exploratory Analysis (level 2)** option from the **Other Settings** menu on the main menu bar. Note that for a three-level model, this option is also available for level-3 predictors.



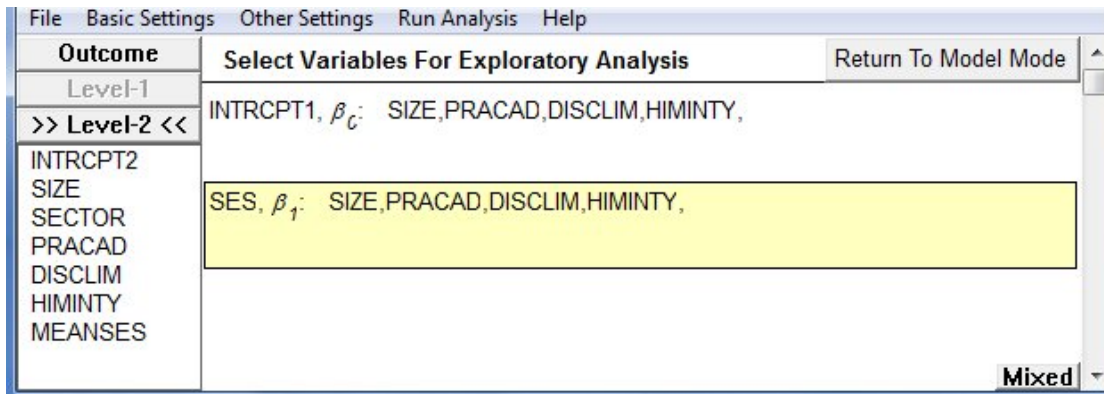
The **Select Variables for Exploratory Analysis** dialog box is displayed.



Click on the variable names at the left to add them to the active equation (in this illustration the slope equation).



Repeat the procedure for the intercept equation after first clicking on it to select it. Once all variables have been entered as shown below, click **Return to Model Mode** and save the command file prior to running it.



3. Interpreting the output

Partial output for this model is shown below.

Exploratory Analysis: estimated level-2 coefficients and their standard errors obtained by regressing EB residuals on level-2 predictors selected for possible inclusion in subsequent HLM runs

Level-1 Coefficient	Potential Level-2 Predictors			
INTRCPT1, β_0				
	SIZE	PRACAD	DISCLIM	HIMINTY
Coefficient	0.000	0.690	-0.161	-0.543
Standard Error	0.000	0.404	0.106	0.229
t-value	1.569	1.707	-1.515	-2.372
SES, β_1				
	SIZE	PRACAD	DISCLIM	HIMINTY
Coefficient	0.000	0.039	-0.005	-0.058
Standard Error	0.000	0.044	0.012	0.025
t-value	1.297	0.899	-0.425	-2.339

The results of this exploratory analysis suggest that HIMINTY might be a good candidate to include in the INTRCPT1 model. The t -values represent the approximate result that will be obtained when one additional predictor is added to any of the level-2 equations. This means that if HIMINTY is added to the model for the INTRCPT1, for example, the apparent relationship suggested above for HIMINTY in the SES slope model might disappear. (For a further discussion of the use of these statistics see discussion in *Hierarchical Linear Models*, p. 270 on “Approximate t -to-Enter Statistics.”)