Scientific Software International

General linear hypothesis testing using the HSB data

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 fourth 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

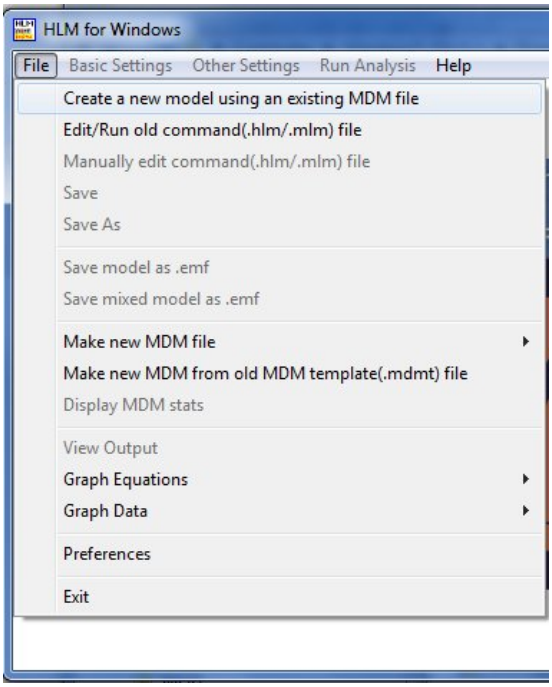
Figure 1.2 First ten cases in HSB2.SAV

2. Creating the command file

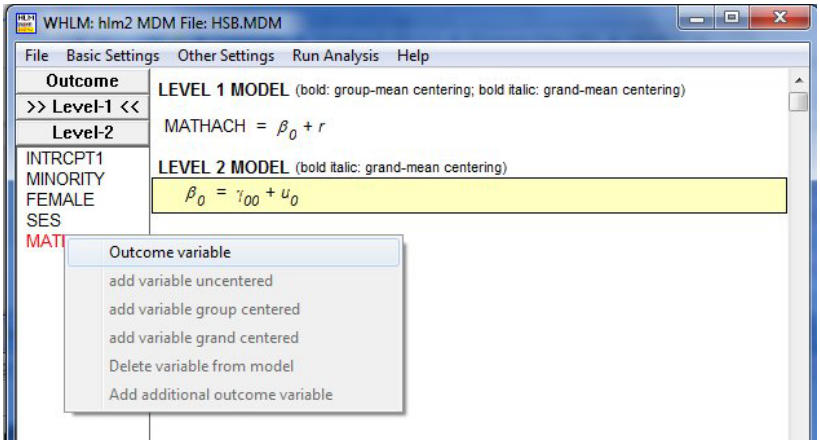
Users may wish to estimate models that allow for heterogeneous level-1 variances. A simple example (see HSB2.HLM) using the HS&B data would be a model that postulates that the two

genders have different means in and variances of math achievement scores. To specify a model that hypothesizes different central tendency and variability in math achievement for the two genders, the model must first be set up, using the existing MDM file.

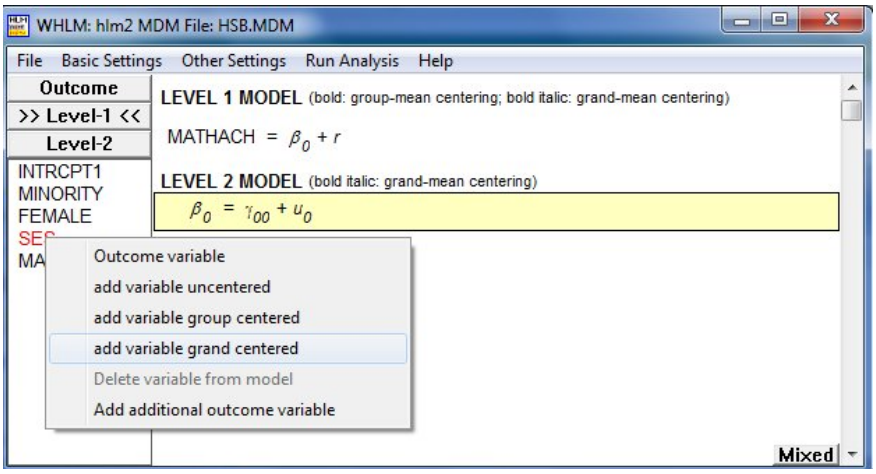
Start by selecting the **Create a new model using an existing MDM file** option from the **File** menu.



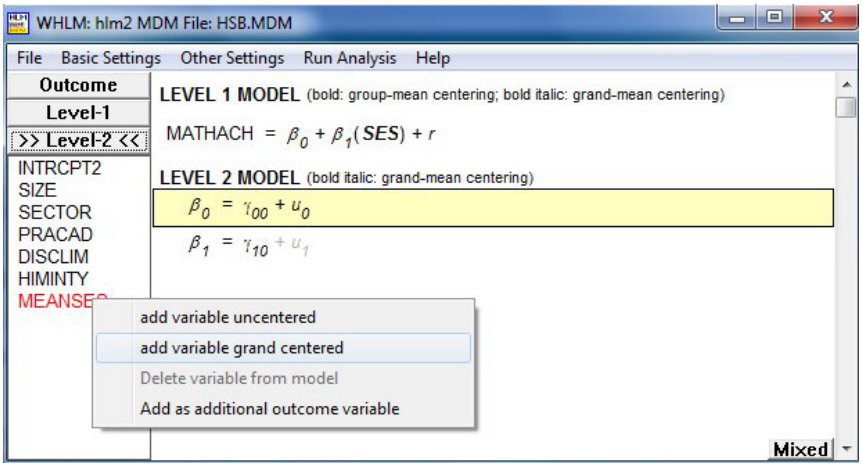
Next, browse for the MDM file **HSB.MDM**. Click **OK** to return to the main window, where the names of the variables contained in the MDM file are now displayed at the left of the window. Select the variable MATHACH as outcome variable, as shown below.



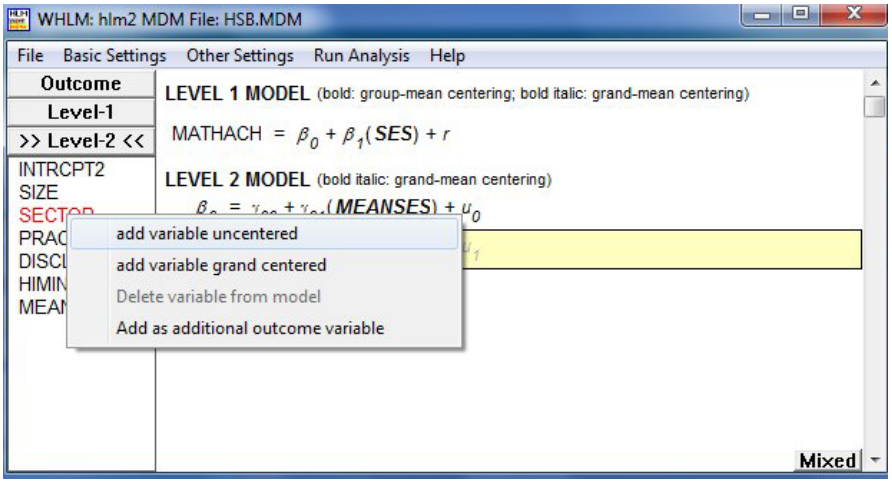
Select the variable SES, representing a student's individual socio-econnmic status, as grand-mean centered level-1 predictor.



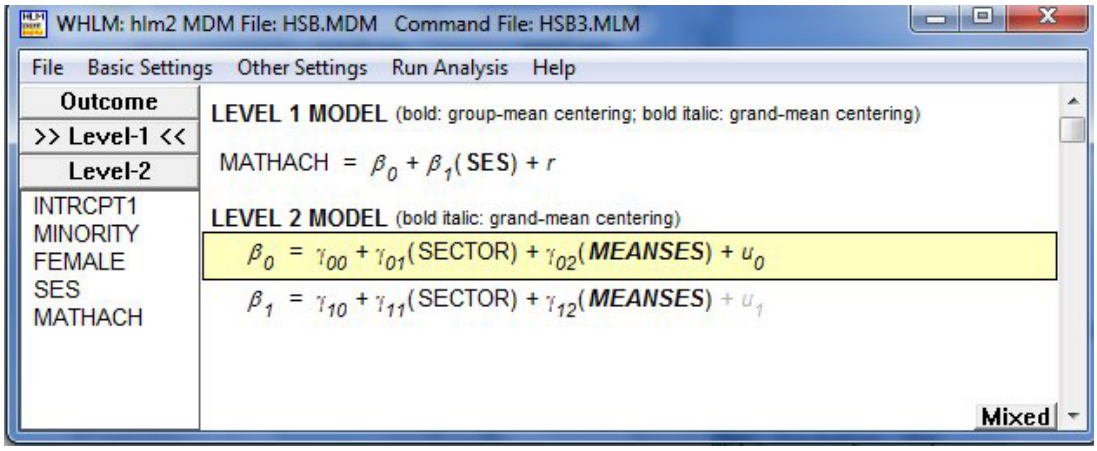
At level-2, first add the predictor MEANSES as grand-mean centered predictor to both the intercept and the slope equation.



Finally, add the level-2 predictor SECTOR as uncentered predictor to both level-2 equations.



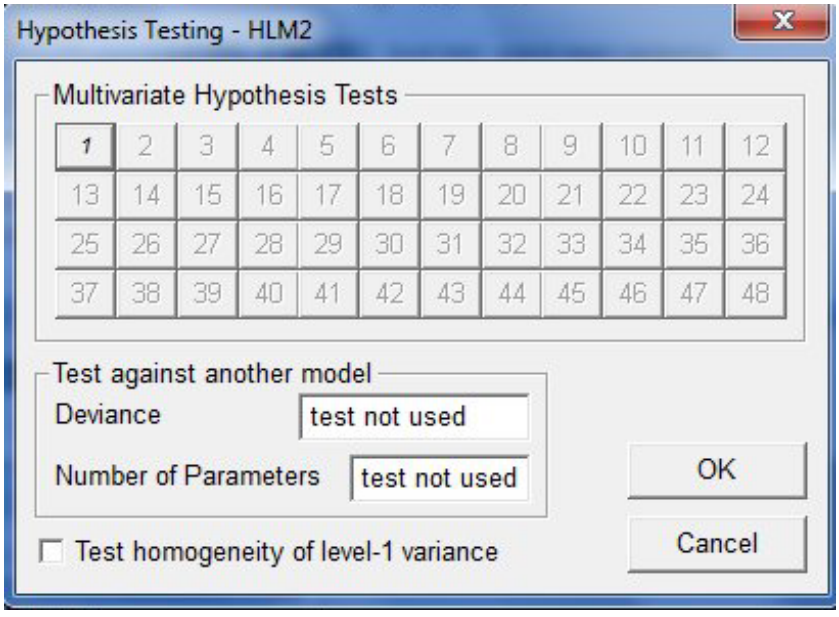
This completes the specification of the basic model of interest.



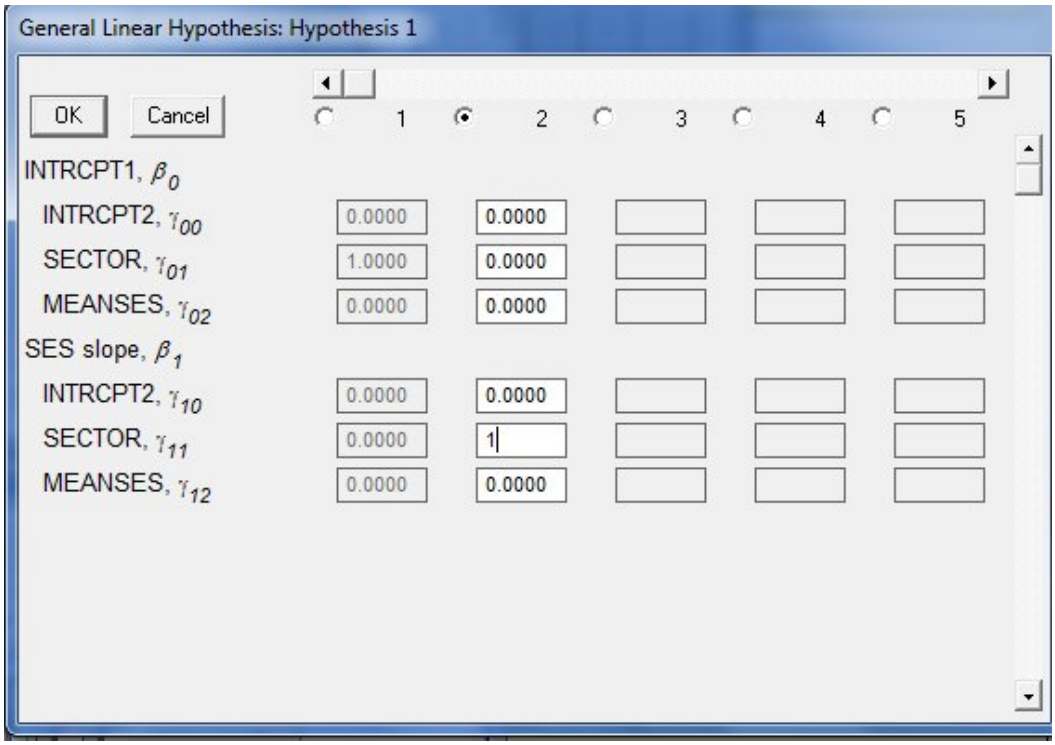
Suppose that it is of interest to test the null hypothesis

$$H_0 : \gamma_{01} = \gamma_{11} = 0$$

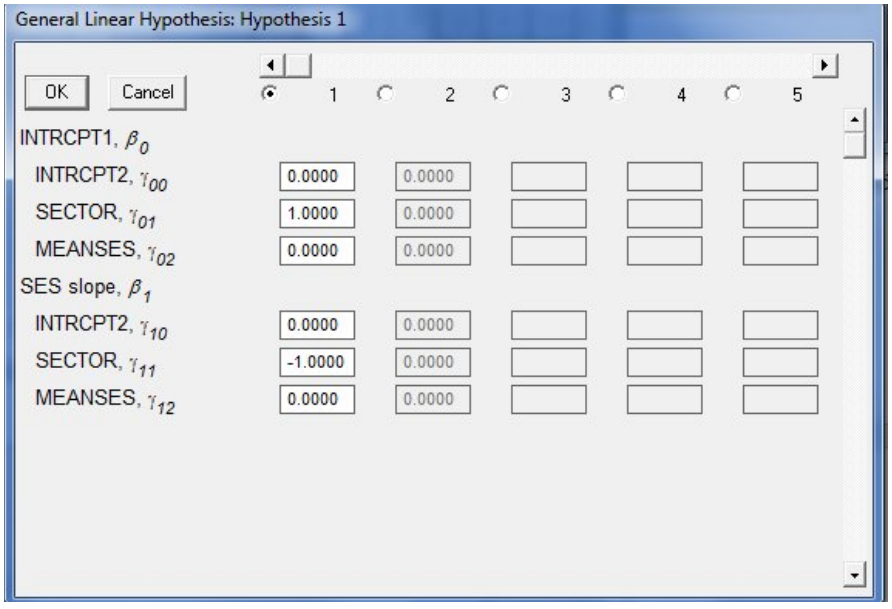
where γ_{01} is the effect of SECTOR on the intercept and γ_{11} is the effect of SECTOR on the SES slope. To set up this hypothesis, select the **Hypothesis Settings** option from the **Other Settings** menu on the main menu bar. The **Hypothesis Testing – HLM2** dialog box is displayed.



Click on "1" to open the **General Linear Hypothesis** dialog box.



In the first column, enter the value "1" in the box next to SECTOR, γ_{01}. Click the radio button at the top to proceed to the second column and enter the value "1" next to SECTOR, γ_{11}. Click **OK** to return to the **Hypothesis Testing** dialog box, and **OK** to return to the main window. Save the model using the **File**, **Save As** option prior to clicking **Run**.



Note that if the hypothesis of interest was

$$H_0 : \gamma_{01} - \gamma_{11} = 0$$

the **General Linear Hypothesis** dialog box setup would have been as shown above.

Once the iterative procedure has converged, the output will automatically be displayed in the format set on the **Preferences** dialog box.

3. Interpreting the output

Partial output for this model is shown below.

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	12.095250	0.173679	69.641	157	<0.001
SECTOR, γ_{01}	1.224401	0.308507	3.969	157	<0.001
MEANSES, γ_{02}	5.336698	0.334617	15.949	157	<0.001
For SES slope, β_1					
INTRCPT2, γ_{10}	2.935664	0.147576	19.893	7022	<0.001
SECTOR, γ_{11}	-1.642102	0.237223	-6.922	7022	<0.001
MEANSES, γ_{12}	1.044120	0.332897	3.136	7022	0.002

Results of General Linear Hypothesis Testing - Test 1

	Coefficients	Contrast	
For INTRCPT1, β_0			
INTRCPT2, γ_{00}	12.095250	0.0000	0.0000
SECTOR, γ_{01}	1.224401	1.0000	0.0000
MEANSES, γ_{02}	5.336698	0.0000	0.0000
For SES slope, β_1			
INTRCPT2, γ_{10}	2.935664	0.0000	0.0000
SECTOR, γ_{11}	-1.642102	0.0000	1.0000
MEANSES, γ_{12}	1.044120	0.0000	0.0000
Estimate		1.2244	-1.6421
Standard error of estimate		0.3085	0.2372

χ^2 statistic = 60.527852
Degrees of freedom = 2
p-value = <0.001

The final results for the fixed effects are followed by the results of the hypothesis test requested. The χ^2-square statistic and associated *p*-value indicate that it is highly unlikely that the observed estimates for γ_{01} and γ_{11} could have occurred under the specified null hypothesis. Recall that the

variable SECTOR represents either Catholic or public schools. From the final table of results for fixed effects we see that while the estimated intercept for a student from a Catholic high school is expected to be 1.22 higher than that of a student from a public school (holding other predictors constant). In contrast, the expected SES slope for the same Catholic school student is expected to be -1.6421 lower than for the public school student (holding all other predictors constant).