

SC708: Hierarchical Linear Modeling
Instructor: Natasha Sarkisian
Class notes: HLM Model Building Strategies

Model Selection Strategy

To summarize, we saw that multilevel models can include 3 types of predictors:

- Level-1 predictors (e.g., student SES)
- Level-2 predictors (e.g., school SECTOR)
- Level-1 predictors aggregated to level 2 (e.g., MEANSES)

In addition, we have a number of choices:

- The intercept can be estimated as either fixed or random (typically random)
- The effects of level 1 predictors can be estimated as either fixed effects or random effects
- Level 2 predictors can be used to predict the intercept (i.e., as direct predictors of DV)
- Level 2 predictors can explain the variation in slopes of level 1 predictors (i.e., as cross-level interactions)

Because so many components are involved, it is best to proceed incrementally and use hypothesis testing. We will first learn how to test hypotheses and what hypotheses we might want to test, and then discuss the main approaches to strategic model building.

Using Hypothesis Testing to Build Models

When making decisions what variables to include and whether to estimate random or fixed effects, we need to use hypothesis testing tools. HLM allows you to test various hypotheses which can be helpful when evaluating which variables and which random effects to include in your model. The basic idea behind hypothesis testing is to build a set of contrasts that would add up to zero under the null hypothesis, and then test the hypothesis that, combined, they are indeed zero. We will do some recodes to HSB data and reimport it for the next example.

```
. recode size (0/499=1) (500/1199=2) (1200/3000=3), gen(sized)
(7185 differences between size and sized)
```

```
. tab sized, gen(sized)
```

RECODE of size	Freq.	Percent	Cum.
1	1,524	21.21	21.21
2	2,984	41.53	62.74
3	2,677	37.26	100.00
Total	7,185	100.00	

```
. tab1 sized1 sized2 sized3
```

```
-> tabulation of sized1
```

sized== 1.0000	Freq.	Percent	Cum.
-------------------	-------	---------	------

0	5,661	78.79	78.79
1	1,524	21.21	100.00

Total	7,185	100.00	

-> tabulation of sized2

sized==	Freq.	Percent	Cum.
2.0000			

0	4,201	58.47	58.47
1	2,984	41.53	100.00

Total	7,185	100.00	

-> tabulation of sized3

sized==	Freq.	Percent	Cum.
3.0000			

0	4,508	62.74	62.74
1	2,677	37.26	100.00

Total	7,185	100.00	

The outcome variable is MATHACH

The model specified for the fixed effects was:

Level-1 Coefficients	Level-2 Predictors
INTRCPT1, B0	INTRCPT2, G00 SECTOR, G01 SIZED1, G02 SIZED3, G03
FEMALE slope, B1	INTRCPT2, G10 SECTOR, G11 SIZED1, G12 SIZED3, G13
SES slope, B2	INTRCPT2, G20 SECTOR, G21 SIZED1, G22 SIZED3, G23

The model specified for the covariance components was:

Sigma squared (constant across level-2 units)

Tau dimensions

INTRCPT1
FEMALE slope
SES slope

Summary of the model specified (in equation format)

Level-1 Model

$$Y = B0 + B1*(FEMALE) + B2*(SES) + R$$

Level-2 Model

$$B0 = G00 + G01*(SECTOR) + G02*(SIZED1) + G03*(SIZED3) + U0$$

$$B1 = G10 + G11*(SECTOR) + G12*(SIZED1) + G13*(SIZED3) + U1$$

$$B2 = G20 + G21*(SECTOR) + G22*(SIZED1) + G23*(SIZED3) + U2$$

Iterations stopped due to small change in likelihood function

***** ITERATION 242 *****

Sigma_squared = 36.36622

Tau

INTRCPT1,B0	4.27594	-0.99430	0.62032
FEMALE,B1	-0.99430	0.89361	-0.15189
SES,B2	0.62032	-0.15189	0.12308

Tau (as correlations)

INTRCPT1,B0	1.000	-0.509	0.855
FEMALE,B1	-0.509	1.000	-0.458
SES,B2	0.855	-0.458	1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.650
FEMALE, B1	0.196
SES, B2	0.060

Note: The reliability estimates reported above are based on only 123 of 160 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

The value of the likelihood function at iteration 242 = -2.325021E+004

The outcome variable is MATHACH

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	12.264554	0.407791	30.076	156	0.000
SECTOR, G01	2.370436	0.443345	5.347	156	0.000
SIZED1, G02	-1.230907	0.577382	-2.132	156	0.034
SIZED3, G03	0.523767	0.475612	1.101	156	0.273
For FEMALE slope, B1					
INTRCPT2, G10	-0.992369	0.360317	-2.754	156	0.007
SECTOR, G11	-0.174713	0.413095	-0.423	156	0.673
SIZED1, G12	0.711859	0.527660	1.349	156	0.179
SIZED3, G13	-0.598786	0.431117	-1.389	156	0.167
For SES slope, B2					
INTRCPT2, G20	2.815012	0.223272	12.608	156	0.000
SECTOR, G21	-1.284821	0.239249	-5.370	156	0.000
SIZED1, G22	0.254340	0.295899	0.860	156	0.392
SIZED3, G23	0.125232	0.259568	0.482	156	0.630

The outcome variable is MATHACH

Final estimation of fixed effects
(with robust standard errors)

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

For INTRCPT1, B0					
INTRCPT2, G00	12.264554	0.406861	30.144	156	0.000
SECTOR, G01	2.370436	0.429390	5.520	156	0.000
SIZED1, G02	-1.230907	0.580527	-2.120	156	0.035
SIZED3, G03	0.523767	0.452241	1.158	156	0.249
For FEMALE slope, B1					
INTRCPT2, G10	-0.992369	0.371180	-2.674	156	0.009
SECTOR, G11	-0.174713	0.417936	-0.418	156	0.676
SIZED1, G12	0.711859	0.505528	1.408	156	0.161
SIZED3, G13	-0.598786	0.422319	-1.418	156	0.158
For SES slope, B2					
INTRCPT2, G20	2.815012	0.209242	13.453	156	0.000
SECTOR, G21	-1.284821	0.225575	-5.696	156	0.000
SIZED1, G22	0.254340	0.286061	0.889	156	0.376
SIZED3, G23	0.125232	0.236842	0.529	156	0.597

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value

INTRCPT1, U0	2.06783	4.27594	119	348.85471	0.000
FEMALE slope, U1	0.94531	0.89361	119	139.73919	0.094
SES slope, U2	0.35083	0.12308	119	129.90176	0.233
level-1, R	6.03044	36.36622			

Note: The chi-square statistics reported above are based on only 123 of 160 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

Statistics for current covariance components model

Deviance = 46500.428022
Number of estimated parameters = 7

1. Single parameter tests of significance.

Single parameter tests are presented in your regular HLM output; in practice, there is no need to run such tests in addition to the regular output, but for learning purposes, we will start with these. Suppose we want to test whether a specific coefficient (e.g. the SES slope for average SES public schools (i.e., intercept for SES slope, gamma 20) is zero. The set of contrasts that we will specify for that will include 1 for G20 and 0 for everything else; therefore, we will test $H_0: G20=0$.

Results of General Linear Hypothesis Testing

	Coefficients	Contrast

For INTRCPT1, B0		

INTRCPT2, G00	12.264554	0.000
SECTOR, G01	2.370436	0.000
SIZED1, G02	-1.230907	0.000
SIZED3, G03	0.523767	0.000
For FEMALE slope, B1		
INTRCPT2, G10	-0.992369	0.000
SECTOR, G11	-0.174713	0.000
SIZED1, G12	0.711859	0.000
SIZED3, G13	-0.598786	0.000
For SES slope, B2		
INTRCPT2, G20	2.815012	1.000
SECTOR, G21	-1.284821	0.000
SIZED1, G22	0.254340	0.000
SIZED3, G23	0.125232	0.000

Chi-square statistic = 180.993971
Degrees of freedom = 1
P-value = 0.000000

Here, we reject Ho.

2. Multi-parameter tests of significance.

Here, we test the hypothesis that multiple coefficients are all equal to 0. Typically, we do that in order to decide whether they can be omitted from the model. This can either be coefficients for different variables (possibly related, e.g. sets of dummies), or coefficients for the same variable in different parts of the model. For example, for could test that all coefficients for SES slope are zero. That would mean testing a combined hypothesis:

G20=0

G21=0

G22=0

G23=0

We can do that by selecting 1 for each of these coefficients when selecting contrasts (separate column for each):

Results of General Linear Hypothesis Testing

		Coefficients				Contrast			

For	INTRCPT1, B0								
	INTRCPT2, G00	12.264554	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SECTOR, G01	2.370436	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SIZED1, G02	-1.230907	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SIZED3, G03	0.523767	0.000	0.000	0.000	0.000	0.000	0.000	0.000
For	FEMALE slope, B1								
	INTRCPT2, G10	-0.992369	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SECTOR, G11	-0.174713	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SIZED1, G12	0.711859	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	SIZED3, G13	-0.598786	0.000	0.000	0.000	0.000	0.000	0.000	0.000
For	SES slope, B2								
	INTRCPT2, G20	2.815012	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	SECTOR, G21	-1.284821	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	SIZED1, G22	0.254340	0.000	0.000	1.000	0.000	0.000	0.000	0.000
	SIZED3, G23	0.125232	0.000	0.000	0.000	1.000	0.000	0.000	0.000

Chi-square statistic = 542.858054
Degrees of freedom = 4
P-value = 0.000000

We reject H_0 ; the coefficients associated with SES slope are jointly significant. We can also test whether SECTOR is significant across equations:

Results of General Linear Hypothesis Testing

		Coefficients		Contrast		
For	INTRCPT1, B0					
	INTRCPT2, G00	12.264554	0.000	0.000	0.000	
	SECTOR, G01	2.370436	1.000	0.000	0.000	
	SIZED1, G02	-1.230907	0.000	0.000	0.000	
	SIZED3, G03	0.523767	0.000	0.000	0.000	
For	FEMALE slope, B1					
	INTRCPT2, G10	-0.992369	0.000	0.000	0.000	
	SECTOR, G11	-0.174713	0.000	1.000	0.000	
	SIZED1, G12	0.711859	0.000	0.000	0.000	
	SIZED3, G13	-0.598786	0.000	0.000	0.000	
For	SES slope, B2					
	INTRCPT2, G20	2.815012	0.000	0.000	0.000	
	SECTOR, G21	-1.284821	0.000	0.000	1.000	
	SIZED1, G22	0.254340	0.000	0.000	0.000	
	SIZED3, G23	0.125232	0.000	0.000	0.000	

Chi-square statistic = 88.796643
 Degrees of freedom = 3
 P-value = 0.000000

This test is also used to jointly test whether multiple variables have non-significant coefficients and therefore can be omitted; e.g., let's test whether everything that is non-significant here can be omitted:

Results of General Linear Hypothesis Testing

		Coefficients		Contrast				
For	INTRCPT1, B0							
	INTRCPT2, G00	12.264554	0.000	0.000	0.000	0.000	0.000	0.000
	SECTOR, G01	2.370436	0.000	0.000	0.000	0.000	0.000	0.000
	SIZED1, G02	-1.230907	0.000	0.000	0.000	0.000	0.000	0.000
	SIZED3, G03	0.523767	1.000	0.000	0.000	0.000	0.000	0.000
For	FEMALE slope, B1							
	INTRCPT2, G10	-0.992369	0.000	0.000	0.000	0.000	0.000	0.000
	SECTOR, G11	-0.174713	0.000	1.000	0.000	0.000	0.000	0.000
	SIZED1, G12	0.711859	0.000	0.000	1.000	0.000	0.000	0.000
	SIZED3, G13	-0.598786	0.000	0.000	0.000	1.000	0.000	0.000
For	SES slope, B2							
	INTRCPT2, G20	2.815012	0.000	0.000	0.000	0.000	0.000	0.000
	SECTOR, G21	-1.284821	0.000	0.000	0.000	0.000	0.000	0.000
	SIZED1, G22	0.254340	0.000	0.000	0.000	0.000	1.000	0.000
	SIZED3, G23	0.125232	0.000	0.000	0.000	0.000	0.000	1.000

Chi-square statistic = 10.149290
 Degrees of freedom = 6
 P-value = 0.117729

3. Tests for equality of coefficients.

We can also test whether two or more coefficients are equal. This is typically used when we have a series of related dummy variables, and we want to combine some dummies. E.g., we could have students' racial/ethnic identification, with dummy variables representing African American, Mexican American, Puerto Rican, Asian American, etc., the omitted category being

White. We could then wonder whether we could simplify that into one dichotomy, White vs ethnic minority. But to test whether the data support this simplification, we'd test whether coefficients for each ethnic group equal to each other (i.e., are all groups different from Whites in the same way?). We have SIZED1 and SIZED3 dummies here, so let's test whether they can be combined. We test:

G02=G03

G12=G13

G22=G23

To do this using contrasts, we redefine it as:

G02-G03=0

G12-G13=0

G22-G23=0

Results of General Linear Hypothesis Testing					
		Coefficients	Contrast		
For	INTRCPT1, B0				
	INTRCPT2, G00	12.264554	0.000	0.000	0.000
	SECTOR, G01	2.370436	0.000	0.000	0.000
	SIZED1, G02	-1.230907	1.000	0.000	0.000
	SIZED3, G03	0.523767	-1.000	0.000	0.000
For	FEMALE slope, B1				
	INTRCPT2, G10	-0.992369	0.000	0.000	0.000
	SECTOR, G11	-0.174713	0.000	0.000	0.000
	SIZED1, G12	0.711859	0.000	1.000	0.000
	SIZED3, G13	-0.598786	0.000	-1.000	0.000
For	SES slope, B2				
	INTRCPT2, G20	2.815012	0.000	0.000	0.000
	SECTOR, G21	-1.284821	0.000	0.000	0.000
	SIZED1, G22	0.254340	0.000	0.000	1.000
	SIZED3, G23	0.125232	0.000	0.000	-1.000

Chi-square statistic = 14.730274

Degrees of freedom = 3

P-value = 0.002467

Here, we reject Ho, so we wouldn't be able to combine those variables.

If we had a set of four dummies (one omitted), we would do pairwise tests:

G02-G03=0

G02-G04=0

G12-G13=0

G12-G14=0

G22-G23=0

G22-G24=0

4. Tests for variance components

If we are interested in testing hypotheses about variance components or their combinations (e.g., see step 3 in both model-building algorithms), we should utilize likelihood ratio tests based on

deviance values. To use such test, we estimate two models, and calculate D0-D1. The resulting difference follows chi-square distribution with df=number of parameters for model 0 minus number of parameters for model 1.

Here, let's see whether we can jointly fix the slopes for FEMALE and SES (but such tests for a single variance component also can be useful because it is a joint test for variance and all covariances). For the full model above, we had:

```
Statistics for current covariance components model
-----
Deviance = 46500.428022
Number of estimated parameters = 7
```

Level-1 Model

$$Y = B_0 + B_1*(FEMALE) + B_2*(SES) + R$$

Level-2 Model

$$\begin{aligned} B_0 &= G_{00} + G_{01}*(SECTOR) + G_{02}*(SIZED1) + G_{03}*(SIZED3) + U_0 \\ B_1 &= G_{10} + G_{11}*(SECTOR) + G_{12}*(SIZED1) + G_{13}*(SIZED3) \\ B_2 &= G_{20} + G_{21}*(SECTOR) + G_{22}*(SIZED1) + G_{23}*(SIZED3) \end{aligned}$$

```
Statistics for current covariance components model
-----
Deviance = 46511.072608
Number of estimated parameters = 2
```

Variance-Covariance components test

```
-----
Chi-square statistic = 10.64459
Number of degrees of freedom = 5
P-value = 0.058
```

P-value indicates that there is no significant difference in model fit between these two models.

Model Development Algorithms

Two main algorithms are recommended; the first one differentiates between level 1 and level 2 variables; the second one does not.

Level-specific algorithm:

1. Fit a fully unconditional model (Model 0). Evaluate level 2 variance to see if HLM is necessary.
2. Estimate a model with random intercept and slopes using only level 1 variables and any necessary interactions among them (Model 2). Make all slopes random, unless you have substantive reasons for separating random and non-random ones. Note, however, that random slopes for interaction terms can be difficult to interpret.
3. Evaluate slope variance, decide whether some slopes should be non-random, and fix those slopes. (Do a joint significance test to doublecheck that all those slopes are jointly not significant.)

4. Based on the significance of regression coefficients, exclude variables where both coefficients and corresponding random effects are not significant. Keep the variable if the coefficient is non-significant but the random effect is. Make sure to conduct hypotheses tests to make sure these variables are jointly not significant. (Note that sometimes you might have substantive reasons to keep the variable even if its coefficient is not significant.)
5. Estimate means-as-outcomes with level 1 covariates model (Model 4) to select level 2 predictors of intercept (include both original level 2 variables and aggregates of level 1). Use hypothesis testing to trim the model.
6. For slopes with significant variance, use level 2 predictors to explain that variance (i.e., estimate an intercepts-and-slopes-as-outcomes model, Model 5). If a slope does not have significant variance but your theory suggests cross-level interaction, do include it. Use hypothesis testing to trim the model.
7. If the slope variance remaining after entering level 2 predictors is not statistically significant, estimate that slope as non-randomly varying (Model 6).

Combined algorithm:

1. Fit a fully unconditional model (Model 0). Evaluate level 2 variance to see if HLM is necessary.
2. Enter all level 2 and level 1 variables in the model, and include any within-level and cross-level interactions based on theory (Model 5). (Don't forget to use aggregates of level 1 variables.) Make all slopes random, unless you have substantive reasons for separating random and non-random ones. Note, however, that random slopes for interaction terms can be difficult to interpret.
3. Evaluate slope variance, decide whether some slopes should be non-random, and fix those slopes (Model 6). (Do a joint significance test to doublecheck that all those slopes are jointly not significant.)
4. Based on the significance of regression coefficients, exclude variables where both coefficients and corresponding random effects are not significant. Keep the variable if the coefficient is non-significant but the random effect is. Make sure to conduct hypotheses tests to make sure these variables are jointly not significant. (Note that sometimes you might have substantive reasons to keep the variable even if its coefficient is not significant.)
5. If there are remaining random slopes with significant variance, consider adding other cross-level interactions to explain that variance. If that leads to the random slope becoming non-significant, estimate that slope as non-randomly varying (Model 6).

The issue of centering

You have already noticed that HLM asks you whether and how you'd like to center your predictors. Here, we will discuss the issues involved in making these decisions.

Level-1 predictors:

1. Natural metric (X):

You should only use the original metric if the value of 0 for a predictor is a meaningful value. When 0 is not meaningful, the estimate of the intercept will be arbitrary and may be estimated

with poor precision. Lack of precision in HLM can be very problematic. First, because you are estimating within-group intercepts, thus with possibly small N, the estimates may be quite unstable. Second, because you may be trying to model variation in these intercepts, your model will be affected by the unreliability of the estimates.

2. Grand-mean centering (X - grand mean):

This will address the problems with estimation of intercept in original metric. Because the 0 values will fall in the middle of the distribution of the predictors, the intercept estimates will be estimated with much more precision. The intercept is also interpretable. Specifically, it will represent the group-mean value for a person with a (grand) average on every predictor. The interpretation of the intercepts is now “adjusted group mean.” The interpretation of slopes does not change. E.g. our measure of SES is already grand-mean centered because it is a standardized scale. So we can interpret the fixed effect for the intercept as the average math achievement adjusted for SES – i.e., the average math achievement for someone with average SES.

LEVEL 1 MODEL

$$MATHACH_{ij} = \beta_{0j} + \beta_{1j}(FEMALE_{ij}) + \beta_{2j}(SES_{ij} - \overline{SES}_{..}) + r_{ij}$$

LEVEL 2 MODEL

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01}(SECTOR_j) + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}(SECTOR_j) + u_{1j} \\ \beta_{2j} &= \gamma_{20} + \gamma_{21}(SECTOR_j) \end{aligned}$$

Level-1 Coefficients	Level-2 Predictors
INTRCPT1, B0	INTRCPT2, G00
	SECTOR, G01
FEMALE slope, B1	INTRCPT2, G10
	SECTOR, G11
# SES slope, B2	INTRCPT2, G20
	SECTOR, G21

'#' - The residual parameter variance for this level-1 coefficient has been set to zero.

Summary of the model specified (in equation format)

Level-1 Model

$$Y = B0 + B1*(FEMALE) + B2*(SES) + R$$

Level-2 Model

$$\begin{aligned} B0 &= G00 + G01*(SECTOR) + U0 \\ B1 &= G10 + G11*(SECTOR) + U1 \\ B2 &= G20 + G21*(SECTOR) \end{aligned}$$

Sigma_squared = 36.45566

Tau

INTRCPT1,B0	4.20999	-1.19503
FEMALE,B1	-1.19503	1.10455

Tau (as correlations)
 INTRCPT1,B0 1.000 -0.554
 FEMALE,B1 -0.554 1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.681
FEMALE, B1	0.236

Final estimation of fixed effects
 (with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	12.437528	0.253308	49.100	158	0.000
SECTOR, G01	2.084354	0.416840	5.000	158	0.000
For FEMALE slope, B1					
INTRCPT2, G10	-1.222739	0.220850	-5.537	158	0.000
SECTOR, G11	0.031679	0.401264	0.079	158	0.938
For SES slope, B2					
INTRCPT2, G20	2.919985	0.141351	20.658	7179	0.000
SECTOR, G21	-1.293137	0.207803	-6.223	7179	0.000

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	2.05183	4.20999	121	347.17034	0.000
FEMALE slope, U1	1.05098	1.10455	121	153.20942	0.025
level-1, R	6.03785	36.45566			

Note that while it may seem inappropriate at first to center a dummy variable, in HLM it actually is quite useful. If uncentered, the intercept in a model with a dummy variable is the average value when the dummy variable is 0. If the dummy variable is centered, the intercept then becomes the mean adjusted for the proportion of cases with the dummy variable=1. For example, if the indicator for sex variable is centered around the grand mean, this centered predictor can take two values. If the subject is female, it will equal the proportion of male students in the sample. If the subject is male, it will equal to minus the proportion of female students in the sample. Zero on this variable becomes the average proportion of female students. The intercept again will be the adjusted group mean – in this case, it is adjusted for the difference among level-2 units in the percentage of female students.

LEVEL 1 MODEL

$$\text{MATHACH}_{ij} = \beta_{0j} + \beta_{1j}(\text{FEMALE}_{ij} - \overline{\text{FEMALE}_{..}}) + \beta_{2j}(\text{SES}_{ij} - \overline{\text{SES}_{..}}) + r_{ij}$$

LEVEL 2 MODEL

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{SECTOR}_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{SECTOR}_j) + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(\text{SECTOR}_j)$$

	Level-1 Coefficients	Level-2 Predictors
	INTRCPT1, B0	INTRCPT2, G00
		SECTOR, G01
%	FEMALE slope, B1	INTRCPT2, G10
		SECTOR, G11
#	SES slope, B2	INTRCPT2, G20
		SECTOR, G21

'#' - The residual parameter variance for this level-1 coefficient has been set to zero.

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

Level-1 Model

$$Y = B0 + B1*(FEMALE) + B2*(SES) + R$$

Level-2 Model

$$B0 = G00 + G01*(SECTOR) + U0$$

$$B1 = G10 + G11*(SECTOR) + U1$$

$$B2 = G20 + G21*(SECTOR)$$

Sigma_squared = 36.45669

Tau

INTRCPT1,B0	3.25698	-0.61167
FEMALE,B1	-0.61167	1.09435

Tau (as correlations)

INTRCPT1,B0	1.000	-0.324
FEMALE,B1	-0.324	1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.774
FEMALE, B1	0.234

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	11.791658	0.214281	55.029	158	0.000
SECTOR, G01	2.101157	0.333510	6.300	158	0.000

For FEMALE slope, B1						
INTRCPT2, G10	-1.222670	0.220851	-5.536	158	0.000	
SECTOR, G11	0.032271	0.401209	0.080	158	0.936	
For SES slope, B2						
INTRCPT2, G20	2.919869	0.141361	20.655	7179	0.000	
SECTOR, G21	-1.292989	0.207806	-6.222	7179	0.000	

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0		1.80471	3.25698	121	488.52692	0.000
FEMALE slope, U1		1.04611	1.09435	121	153.19922	0.025
level-1, R		6.03794	36.45669			

3. Group-mean centering (X – group mean):

Predictors can also be centered around the mean value for the group to which they belong. The intercept can then be interpreted as the average outcome for each group. This allows interpretation of parameter estimates as person-level effects within each group (i.e. if you differ from your group's average by one unit, your math achievement will increase by X units).

Again, we can group-mean center dummy variables as well. For females, we will get a value equal to the proportion of male students in school j; for males, it will take the value equal to minus the proportion of females in that school. The fact that it is a dummy variable does not change the interpretation of the intercept when group mean-centering is employed.

LEVEL 1 MODEL

$$\text{MATHACH}_{ij} = \beta_{0j} + \beta_{1j}(\text{FEMALE}_{ij} - \overline{\text{FEMALE}_{.j}}) + \beta_{2j}(\text{SES}_{ij} - \overline{\text{SES}_{.j}}) + r_{ij}$$

LEVEL 2 MODEL

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{SECTOR}_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{SECTOR}_j) + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(\text{SECTOR}_j)$$

The outcome variable is MATHACH
The model specified for the fixed effects was:

Level-1 Coefficients	Level-2 Predictors
INTRCPT1, B0	INTRCPT2, G00
	SECTOR, G01
* FEMALE slope, B1	INTRCPT2, G10
	SECTOR, G11
#* SES slope, B2	INTRCPT2, G20
	SECTOR, G21

'#' - The residual parameter variance for this level-1 coefficient has been set to zero.

'*' - This level-1 predictor has been centered around its group mean.

Summary of the model specified (in equation format)

Level-1 Model

$$Y = B0 + B1*(FEMALE) + B2*(SES) + R$$

Level-2 Model

$$B0 = G00 + G01*(SECTOR) + U0$$

$$B1 = G10 + G11*(SECTOR) + U1$$

$$B2 = G20 + G21*(SECTOR)$$

Sigma_squared = 36.45732

Tau

INTRCPT1,B0	6.75745	-0.63530
FEMALE,B1	-0.63530	0.82580

Tau (as correlations)

INTRCPT1,B0	1.000	-0.269
FEMALE,B1	-0.269	1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.882
FEMALE, B1	0.188

Note: The reliability estimates reported above are based on only 123 of 160 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

The value of the likelihood function at iteration 31 = -2.330178E+004

The outcome variable is MATHACH
Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	11.393469	0.292627	38.935	158	0.000
SECTOR, G01	2.804207	0.436272	6.428	158	0.000
For FEMALE slope, B1					
INTRCPT2, G10	-1.224963	0.218270	-5.612	158	0.000
SECTOR, G11	0.421184	0.422651	0.997	158	0.321
For SES slope, B2					
INTRCPT2, G20	2.732981	0.156703	17.440	7179	0.000
SECTOR, G21	-1.310898	0.229605	-5.709	7179	0.000

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	2.59951	6.75745	121	890.99031	0.000
FEMALE slope, U1	0.90873	0.82580	121	150.58868	0.035
level-1, R	6.03799	36.45732			

Note: The chi-square statistics reported above are based on only 123 of 160 units that had sufficient data for computation. Fixed effects and variance

components are based on all the data.

Important:

Under grand-mean centering or no centering, the parameter estimates reflect a combination of (1) person-level effects and (2) compositional effects. But when we use a group-centered predictor, we only estimate the person-level effects.

In order not to discard the compositional effects with group-mean centering, level-2 variables should be created to represent the group mean values for each group-mean centered predictor. Because the group mean is effectively removed from the individual scores, the level-2 values will be orthogonal to the level-1 values. Note that while HLM6 has an option for group-mean centering, it does not compute the group mean values of a predictor to be included as a level-2 variable – you have to do that in another statistical package and then import the data into HLM (more on this below).

E.g. we can use group mean centering for SES and using mean SES as a school level variable (here, MEANSES is already in the dataset):

LEVEL 1 MODEL

$$\text{MATHACH}_{ij} = \beta_{0j} + \beta_{1j}(\text{SES}_{ij} - \overline{\text{SES}}_{..}) + r_{ij}$$

LEVEL 2 MODEL

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{SECTOR}_j) + \gamma_{02}(\text{MEANSES}_j) + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(\text{SECTOR}_j) + \gamma_{12}(\text{MEANSES}_j) + u_{1j}$$

```
-----
Level-1                               Level-2
Coefficients                           Predictors
-----
          INTRCPT1, B0                 INTRCPT2, G00
                                         SECTOR, G01
                                         MEANSES, G02
*      SES slope, B1                   INTRCPT2, G10
                                         SECTOR, G11
                                         MEANSES, G12

'' - This level-1 predictor has been centered around its group mean.
Summary of the model specified (in equation format)
-----
```

Level-1 Model

$$Y = B0 + B1*(SES) + R$$

Level-2 Model

$$B0 = G00 + G01*(SECTOR) + G02*(MEANSES) + U0$$

$$B1 = G10 + G11*(SECTOR) + G12*(MEANSES) + U1$$

Sigma_squared = 36.70313

Tau		
INTRCPT1,B0	2.37996	0.19058
SES,B1	0.19058	0.14892

Tau (as correlations)
 INTRCPT1,B0 1.000 0.320
 SES,B1 0.320 1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.733
SES, B1	0.073

Final estimation of fixed effects
 (with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	12.096006	0.173699	69.638	157	0.000
SECTOR, G01	1.226384	0.308484	3.976	157	0.000
MEANSES, G02	5.333056	0.334600	15.939	157	0.000
For SES slope, B1					
INTRCPT2, G10	2.937981	0.147620	19.902	157	0.000
SECTOR, G11	-1.640954	0.237401	-6.912	157	0.000
MEANSES, G12	1.034427	0.332785	3.108	157	0.003

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	1.54271	2.37996	157	605.29503	0.000
SES slope, U1	0.38590	0.14892	157	162.30867	0.369
level-1, R	6.05831	36.70313			

Here, the effects of SES turn out to be quite complex: For those who are in a public school whose SES is at their school's average and whose school itself is average in terms of its SES, the math achievement is 12.096. If you are in a Catholic school with such properties, it's $12.1+1.2=13.3$. But if your school's average SES is 1 unit higher than the average for all schools, then your math achievement increases by 5.33. Further, in addition to these school-level effects, your individual SES also plays a role – if you are in an average (in terms of SES) public school, one unit increase in your SES will raise your math score by 2.94. In a Catholic school, that effect would be $2.94-1.64=1.30$. But if you are in a public school and your school is 1 unit above an average school in its SES, then your personal SES impact (per one unit) would be $2.94+1.03=3.97$. For a Catholic school in that situation, that effect of SES would become $2.94-1.64+1.03=2.33$. Interestingly, personal SES seems to have stronger impact on math achievement in those schools that have relatively high school-level SES.

The choice between grand-mean centering and group-mean centering depends on your theoretical thinking about processes. If you think that the absolute values of level 1 variable matter, then use grand-mean centering. If you think that it is the relative position of the person with regards to their group's mean is what matters, then use group-centering.

Level-2 predictors:

Centering issues for level-2 predictors are essentially the same issues faced in any regression. If the value of 0 for a predictor is not meaningful, the intercept will not have a meaningful interpretation and the estimate may lack precision. When these conditions exist, centering is advisable. You can either use grand-mean centering (then the intercept will reflect the average group) or center around some constant (then the intercept will reflect a group with the value of the predictor equal to that constant). Note that HLM6 only has the grand-mean centering option for level-2 predictors – if you want to center around some other value, you would have to generate such a centered variable in another statistical program and then import the data into HLM. Typically, however, grand-mean centering is just fine.

LEVEL 1 MODEL

$$\text{MATHACH}_{ij} = \beta_{0j} + \beta_{1j}(\text{SES}_{ij} - \overline{\text{SES}}_{..}) + r_{ij}$$

LEVEL 2 MODEL

$$\begin{aligned}\beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{SECTOR}_j - \overline{\text{SECTOR}}_{.}) + \\ &\quad \gamma_{02}(\text{MEANSES}_j - \overline{\text{MEANSES}}_{.}) + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}(\text{SECTOR}_j - \overline{\text{SECTOR}}_{.}) + \\ &\quad \gamma_{12}(\text{MEANSES}_j - \overline{\text{MEANSES}}_{.}) + u_{1j}\end{aligned}$$

Level-1 Coefficients	Level-2 Predictors
-----	-----
INTRCPT1, B0	INTRCPT2, G00
\$	SECTOR, G01
\$	MEANSES, G02
* SES slope, B1	INTRCPT2, G10
\$	SECTOR, G11
\$	MEANSES, G12

'*' - This level-1 predictor has been centered around its group mean.
'\$' - This level-2 predictor has been centered around its grand mean.

Summary of the model specified (in equation format)

Level-1 Model

$$Y = B0 + B1*(SES) + R$$

Level-2 Model

$$\begin{aligned}B0 &= G00 + G01*(SECTOR) + G02*(MEANSES) + U0 \\ B1 &= G10 + G11*(SECTOR) + G12*(MEANSES) + U1\end{aligned}$$

Sigma_squared = 36.70313

Tau

INTRCPT1,B0	2.37996	0.19058
SES,B1	0.19058	0.14892

Tau (as correlations)

INTRCPT1,B0	1.000	0.320
-------------	-------	-------

SES,B1 0.320 1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.733
SES, B1	0.073

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	12.631549	0.140082	90.173	157	0.000
SECTOR, G01	1.226384	0.308484	3.976	157	0.000
MEANSES, G02	5.333056	0.334600	15.939	157	0.000
For SES slope, B1					
INTRCPT2, G10	2.219870	0.108224	20.512	157	0.000
SECTOR, G11	-1.640954	0.237401	-6.912	157	0.000
MEANSES, G12	1.034427	0.332785	3.108	157	0.003

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	1.54271	2.37996	157	605.29503	0.000
SES slope, U1	0.38590	0.14892	157	162.30867	0.369
level-1, R	6.05831	36.70313			

Creating Aggregated Variables from Level 1 Data

In Stata:

You can either use level 1 data file or (if you already have it) a single file for two levels. First sort this file by school id:

```
. use "C:\Documents and Settings\SARKISIN\My Documents\hsb1.dta", clear
```

Use egen command to generate an aggregated variable:

```
. bysort id: egen meanses2=mean(ses)
```

If you have two separate files, you'll end up generating this variable in level 1 file, and then you'll have to create a combined file for two levels by merging the two files:

```
. merge m:1 id using "C:\Documents and Settings\SARKISIN\My Documents\hsb2.dta"
```

```
Result          # of obs.
-----
not matched          0
matched             7,185  (_merge==3)
```

```
. drop _merge
```

In SPSS:

Here you can use command Aggregate to generate a new aggregated variable in level-1 file or a merged file. If you have two separate files, you'll end up generating that variable in level 1 file, and you'll have to transfer it into level 2 (utilizing Merge function in SPSS).

```
AGGREGATE  
/BREAK = id  
/meanses2 = MEAN(ses).
```

Then merge:

```
MATCH FILES /FILE=*  
/TABLE='C:\ HSB2.SAV'  
/BY id.  
EXECUTE.
```