HLM

Hierarchical Linear Modeling

Lesson Two

Lesson Two Plan

Multivariate Multilevel Model

I. The Two-Level Multivariate Model

II. Examining Residuals

III. More Practice in Running HLM

I. The Two-Level Multivariate Model

We are in a position now to extend the simple linear model to the multivariate case:

$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + \varepsilon_{ij}$ within groups

$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + U_{0j}$

between groups

$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + U_{1j}$

With this model we can have a number of X variables in the within-group model, and a number of Z variables in the between-group model. We do not have to model the same Z variables on each of the β's.

This general model can be used to address a number of questions in the social sciences. For example, nearly all research on school effectiveness has been directed at answering four principal questions:

- To what extent do school vary in their outcomes?
- To what extent do outcomes vary for pupils of differing status?
- What school policies and practices improve levels of schooling outcomes?
- What school policies and practices reduce inequalities in outcomes between high- and low-status groups?

The first two questions concern *quality* and *equity*; the last two concern their causes. The same set of questions can be asked of other organizational units, such as the school district or classroom.

After we have collected our data we would like to test the adequacy of our model. How well does it fit the data? *I think about the adequacy of a model as follows:*

a. are the within-group variables statistically significant?

b. do the within-group parameters vary across groups?

b. if they do, can the variance of the parameters be explained by between-group factors?

Consider each of these separately

a. are the within unit variables statistically significant?

- be guided by theory; do not be a slave to the .05 rule. (based on a t-test)

- consider the 'role' of each variable:
  - effects of interest (e.g. race, gender)
  - interaction effects
  - control variables
  - to increase precision
  - sometimes the average effect of a variable is not significant, but its effect varies across groups.

b. do the within-unit parameters vary across groups ?

- examine first a model that has no between-group parameter:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + r_{ij} \qquad (1)$$

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + U_{0j} \\ \beta_{1j} &= \gamma_{10} + U_{1j} \\ \beta_{2j} &= \gamma_{20} + U_{2j} \end{aligned} \qquad (2)$$

substituting (2) into (1) yields:

$$Y_{ij} = (\gamma_{00} + U_{0j}) + (\gamma_{10} + U_{1j})X_{1ij} + (\gamma_{20} + U_{2j})X_{2ij} + r_{ij}$$

Notice the model has both individual level residuals r_{ij} and group-level residuals U_{0j}, U_{1j}, U_{2j}.

What we would normally think of as the 'intercept' is now made up of two components - a fixed effect, γ_{00}, and a random effect U_{0j}, which varies across schools.

HLM provides a t-test of the hypothesis that the fixed effect is zero (H_0: $\gamma_{00} = 0$), and a χ^2 test that the variance in U_{0j} is zero [H_0: $Var(U_{0j}) = 0$].

Similarly, the slope for X_1 is made up of a fixed effect , γ_{10} and a random effect U_{1j}. One can think of the slope as an average slope γ_{10} and an increment that is unique to each school, U_{1j}. Again we ask whether the average slope is significant, and whether the schools vary in their slopes.

The same applies for X_2.

If the variances of the group-level residuals are insignificant, we constrain them to be zero. Notice that in equation 2, 💣's are constants, and thus we can write:

$$\mathrm{Var}(👓_{0j}) = \mathrm{Var}(U_{0j}) \quad \mathrm{Var}(👓_{1j}) = \mathrm{Var}(U_{1j})$$

$$\mathrm{Var}(👓_{2j}) = \mathrm{Var}(U_{2j})$$

In the language of HLM, these are called parameter variances. If we can explain the parameter variance with some group-level variables, the variance of the residuals decreases. We then call them residual parameter variances.

c. If the parameter variances are non-zero, it means we have some between-group variance to work with for examining the effects of some group-level variables. Once these are in the model they play a similar role to the fixed portion of the within-group effects.