Multivariate Multilevel Models

- Definition
- Purpose
- Model Structure
- Set-Up and Estimation (SPSS & HLM)
- Interpretation of Results

Definition

- **Multivariate**: Multiple outcomes are modelled simultaneously, e.g. reading, math and science scores

- **Multilevel**: These outcomes are regressed on variables measured at different levels, e.g. student level and school level

Like for growth models, the use of a multivariate model
(as opposed to a univariate model)
requires an extra level:

the Intra-Individual Level.

Purpose

- The multivariate model allows estimation of the correlation between pairs of outcomes at both levels.
- For example, if the two outcomes of interest are math and reading, the multivariate model can determine whether students who have high math scores also have high reading scores, and similarly, whether schools with high mean reading achievement also have high mean mathematics achievement.

A model *could* be estimated separately for each outcome…

Reading:

$Y_{rij} = \beta_{r0j} + r_{rij}$ Level 1

$\beta_{r0j} = \gamma_{r00} + U_{r0j}$ Level 2

Math:

$Y_{mij} = \beta_{m0j} + r_{mij}$ Level 1

$\beta_{m0j} = \gamma_{m00} + U_{m0j}$ Level 2

However, estimating these two models separately does not let us find out the adjusted and unadjusted correlation between r_{rij} and r_{mij} (student level) , and between U_{rij} and U_{mij} (school level) .

Model Structure

Level 1: Intra-Student

The data sheet looks as follows:

STUDNTID	SCHOOLID	SCORE	DREAD	DMATH	DSCIENCE
151	15	455	1	0	0
151	15	475	0	1	0
151	15	440	0	0	1
152	15	380	1	0	0
152	15	395	0	1	0
152	15	385	0	0	1

Score = Y_{rij}*DREAD + Y_{mij}*DMATH + Y_{sij}*DSCIENCE

where DREAD, DMATH and DSCIENCE are dummy variables (also called indicators in this context) of whether a given score is for reading, mathematics or science, and where the Y's are coefficients. Note that there is no random term.

Looking back to the previous table, and "estimating" the model for each case, we get:

$455 = \mathbf{Y_{r1j}}$ **Student 1**

$475 = \mathbf{Y_{m1j}}$

$440 = \mathbf{Y_{s1j}}$

$380 = \mathbf{Y_{r2j}}$ **Student 2**

$395 = \mathbf{Y_{m2j}}$

$385 = \mathbf{Y_{s2j}}$

This means that the estimate of the $\mathbf{Y_{rij}}$ coefficient for each student is simply his/her reading score.

$\mathbf{Y_{rij}}$
455
380
(…)

Note that this is the same reading/outcome variable that we were using to estimate the univariate two-level model.

And similarly for the $\mathbf{Y}_{\mathbf{mij}}$ and $\mathbf{Y}_{\mathbf{sij}}$ coefficients.

Therefore, the purpose of the Level 1 model is simply to partition out the different types of score/outcome.

Level 2: Student

$Y_{rij} = \beta_{r0j} + \beta_{r1j}X_{1ij} + r_{rij}$ Reading

$Y_{mij} = \beta_{m0j} + \beta_{m1j}X_{1ij} + r_{mij}$ Math

$Y_{sij} = \beta_{s0j} + \beta_{s1j}X_{1ij} + r_{sij}$ Science

The student-level equation for the multivariate model is thus the same as the student-level equation for the univariate model, except that there is one equation per outcome. Note that each outcome need not be regressed on the same covariates. Moreover, it is now possible to examine the relationship between the student-level residuals.

Level 3: School

$$\beta_{r0j} = \gamma_{r00} + \gamma_{r01}Z_{1j} + \gamma_{r02}Z_{2j} + U_{0j}$$ Reading

$$\beta_{r1j} = \gamma_{r10} + \gamma_{r11}Z_{1j} + \gamma_{r02}Z_{2j} + U_{1j}$$

$$\beta_{m0j} = \gamma_{m00} + \gamma_{m01}Z_{1j} + \gamma_{m02}Z_{2j} + U_{0j}$$ Math

$$\beta_{m1j} = \gamma_{m10} + \gamma_{m11}Z_{1j} + \gamma_{m02}Z_{2j} + U_{1j}$$

$$\beta_{s0j} = \gamma_{s00} + \gamma_{s01}Z_{1j} + \gamma_{s02}Z_{2j} + U_{0j}$$ Science

$$\beta_{s1j} = \gamma_{s10} + \gamma_{s11}Z_{1j} + \gamma_{s02}Z_{2j} + U_{1j}$$

The school-level equations are as before, except that there is a model for each outcome. Again, the school-level covariates do not need to be the same for all outcomes, and it is now possible to examine the relationship between the school-level residuals.