HLM: A Gentle Introduction

D. Betsy McCoach, Ph.D.
Associate Professor, Educational Psychology
Neag School of Education
University of Connecticut

Other names for HLMs

• “Multilevel Linear Models” incorporate higher level predictors into the analysis.
• Raudenbush and Bryk (2002) call these “HLMs”
  – Intercepts and slopes as outcomes
    • Trying to model or explain variability in level one intercepts and slopes, based on group level characteristics
• Also referred to as contextual models, mixed-effects models, random-effects models, variance components models
• The overall model is considered to be a linear model.
Clustering

• Research has consistently demonstrated that people within a particular group or context tend to be more similar to each other in terms of an outcome variable than they are to people in a different group or context.
  – Statistically, we need to use techniques that consider this dependence of the outcome variable between people from the same context

• Repeated observations share this same phenomena (correlated observations)

Nesting Problem-- Non-independence

• Students nested within classrooms or schools are typically more homogeneous than observations from a non-nested study.

• This correlation or dependency (ICC or rho=ρ) violates the assumption of independence necessary for traditional statistics.

• Even mild violations can lead to severe problems with Type I error (typically inflated).
  – Kish (1965): Cluster effect
Intra-class correlation

• Tendency for values of a variable within a cluster to be correlated among themselves relative to the values outside the cluster

• The average correlation between variables within the same higher level unit will be higher than the average correlation between variables for students from different higher-level units

How big an ICC is problematic?

- the degree to which the standard error is affected depends on both the ICC and the number of people per cluster
Design Effect: DEFT

\[ d = \sqrt{1 + (n_j - 1) \cdot \rho} \]

- With clustered data, the ICC is positive and the standard error (which assumes SRS) is underestimated.
- DEFT shows how much the standard error needs to increase as we move from a simple random sample to a cluster sample of size \( n \).
- Multiply SE by the DEFT.

DEFT

\[ d = \sqrt{1 + (n_j - 1) \cdot \rho} \]

- The bigger the cluster, the bigger the DEFT.
- The bigger the ICC, the bigger the DEFT.
- But even small ICC’s can have large consequences.
- The DEFT for an ICC of 0.05 with a cluster size of 100 is

\[ 2.44 = \sqrt{1 + (100 - 1) \cdot 0.05} \]
DEFT as a function of Nj
(ICC held constant at .15)

DEFT as a function of the ICC
(Nj held constant at 25)
Historical Approaches

1. Disaggregate the data (students from same school get same value for the macro level variable)
   – Violates assumption of independence.
   – Inflated type I errors.
2. Aggregate the data (use school means as the outcome)
   – Not measuring the same thing once aggregated
   – Analyzing group data, but formulating conclusions as if were at the individual level!
   – Robinson effect, also called Ecological fallacy

HLM Solution

Consider effects at both levels simultaneously with a level-1 and a level-2 model!
Multilevel Research

• A multilevel data structure allows for an understanding of how person-level variables (e.g., gender, behavior, attitudes, knowledge, general health) as well as group-level variables (e.g., characteristics of classrooms or teachers, neighborhoods, families) can be used to explain an outcome of interest (measured at the lowest level of the analysis)

• A hierarchical system, with interactions between the individual and group (or “context”) characteristics

• Can also reframe longitudinal models as multilevel, but for today we will stick with organizational models

Reasons for using HLM...

• Addresses the “unit of analysis” question
• Can resolve aggregation bias issues
• Enhances precision of estimates over non-hierarchical methods
• More powerful models for individual growth
  – Individual growth curves
• Allows us to model variability across contexts
• Allows us to partition variance across the levels of analysis
Modeling the Dependency

• Accounting for homogeneity within groups allows for an understanding of how group-level effects could impact the outcome of interest
  – Do group level variables moderate the effect of an independent variable on Y?
  • Conditional Models
    – Cross-level effects, aka cross-level interactions

• HLM provides a methodology for connecting the two (or three, etc.) levels together.

Two-level structure

Children, nested within schools
Three-level structure

When groups of schools belong to certain districts:

District 1

\[ S_1 \quad S_2 \quad S_3 \quad \ldots \quad S_j \]

District 2

District K

25 schools: \[ \hat{Y} = b_0 + b_1 SES \]

Public are blue (0)

Private are red (1)
How are public and private schools different?

\[ \hat{Y} = b_0 + b_1 SES \]

Intercept: The “level” is higher in the private schools.

Slope: The slope is flatter in private schools. The slope is steeper in public schools.

This means that the relationship between SES and math achievement is STRONGER in public schools (where the slope is steeper) and WEAKER in private schools (where slope is flatter). A completely flat slope would indicate NO relationship between SES and math achievement.

Organizational Context

- Ask questions about how effects at level 2 affect the outcome of interest
- Graphs— illustrate how context – the public versus private school distinction – can be important to understanding differences in the relationships between SES and Math Achievement
Level 1: HSB

- 160 level-1 equations, represented by:

\[ Y_{ij} = \beta_{0j} + \beta_{1j}(SES_{ij}) + r_{ij} \]

- For each school, the intercept represents the best estimate for prediction of math achievement, when SES is 0
- For each school, slope represents the effect of SES on math achievement
- For each school, the error term \( r_{ij} \) represents deviation for each student from the fitted model

\[(Y - \hat{Y})\]

Level 2: HSB

- 2 level-2 equations
  - One equation models the variability/differences in the 160 intercepts
  - One equation models the variability/differences in the 160 SES slopes

\[ \beta_{0j} = \gamma_{00} + \gamma_{01}(Sector)_{1j} + u_{0j} \]
\[ \beta_{1j} = \gamma_{10} + \gamma_{11}(Sector)_{1j} + u_{1j} \]
Tau Matrix (example: 2x2)

$T$ contains variances and covariances for the (randomly varying) intercepts and slopes.

$$T = \begin{bmatrix} 
\text{var}(u_{0j}) & \text{cov}(u_{0j}, u_{1j}) \\
\text{cov}(u_{0j}, u_{1j}) & \text{var}(u_{1j}) 
\end{bmatrix} = 
\begin{bmatrix} 
\tau_{00} & \tau_{01} \\
\tau_{10} & \tau_{11} 
\end{bmatrix}$$

Combining terms...

$$Y_{ij} = \gamma_{00} + \gamma_{10} \text{SES}_{ij} + \gamma_{01} \text{Sector}_j + \gamma_{11} \text{SES}_{ij} \text{Sector}_j + u_{0j} + u_{1j} \text{SES}_{ij} + r_{ij}$$

*Our goals for today – how do we develop and interpret this model?*
“High School and Beyond” Data

- n = 7185 students, nested within J = 160 schools (approx. 45 students each school).
- Y = mathematics achievement
- X = student (family) SES
- W_{1} = sector
  - Public school (0)
  - Catholic school (1)

Models we will build...

Our examples (printouts) treat SES as grand-mean centered

- Model 1: One-way Random Effects ANOVA model- Null model
- Model 2: Random Coefficients Model (Includes SES at level 1)
- Model 3: Contextual Model (Includes Sector at level 2)
One-way ANOVA with Random Effects

Level 1: \( Y_{ij} = \beta_{0j} + r_{ij} \)

Level 2: \( \beta_{0j} = \gamma_{00} + u_{0j} \)

Combined Model:
\[ Y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \]

Partitioning variance....

\[ Var Y_{ij} = Var (u_{0j} + r_{ij}) = \tau_{00} + \sigma^2 \]

- The total variability in the dependent variable can be separated into two pieces: that which lies between clusters (\( \tau_{00} \)) and that which is within clusters \( \sigma^2 \)
Intra-class correlation

\[ ICC = \frac{Var(u_{0j})}{Var(u_{0j} + r_{ij})} = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \]

- The proportion of the variance in the outcome that is between the level-2 units.
- Proportion of variance explained by the grouping/clustering structure (Hox, 2002)
- Expected correlation between two randomly chosen units within the same cluster (Hox, 2002)

Intra-class correlation

- Tendency for values of a variable within a cluster to be correlated among themselves relative to the values outside the cluster
- The average correlation between variables within the same higher level unit will be higher than the average correlation between variables for students from different higher-level units
Random Coefficients Model

Level 1

\[ Y_{ij} = \beta_{0j} + \beta_{1j} (SES_{ij}) + r_{ij} \]

Level 2

\[ \beta_{0j} = \gamma_{00} + u_{0j} \]
\[ \beta_{1j} = \gamma_{10} + u_{1j} \]

Combined

\[ Y_{ij} = \gamma_{00} + \gamma_{10} (X_{ij}) + u_{0j} + u_{1j} (X_{ij}) + r_{ij} \]

Random Coefficients Regression Models

Unconditional level-2 model: No predictors at level-2

Both intercepts and slopes are allowed to vary across groups

- Variance-covariance components:
  - intercept variance
  - slope variance
  - intercept-slope covariance
- Level one variance
  - Person-level, \( r_{ij} \)
Characterizing Variability

Assuming a random coefficients model (no W’s, and here with grand mean centering):
\[ \beta_{0j} = \text{intercept (mean outcome)} \text{ a student of average SES for the } j^{th} \text{ school} \]
\[ \beta_{1j} = \text{overall effect of SES within the } j^{th} \text{ school} \]

If Y is achievement:
- Variability in the intercepts reflects variability in achievement across schools holding SES constant at 0 \((\tau_{00})\)
- Variability in slopes reflects differences in effect of SES across schools \((\tau_{qq})\)

Each school described by \((\beta_{0j}, \beta_{1j}, \beta_{2j}, \ldots \beta_{qj})\)

Slope and Intercept Variability

For random coefficients model (no level-2 predictors)

\[ E(\beta_{0j}) = \gamma_{00} \text{ and } Var(\beta_{0j}) = \tau_{00} \]
\[ E(\beta_{qj}) = \gamma_{qj} \text{ and } Var(\beta_{qj}) = \tau_{qq} \]
\[ Cov(\beta_{0j}, \beta_{qj}) = \tau_{0q} \]
- There are q+1 variances, and as many covariances as there are pairs of effects in the model.
- A positive covariance means that schools with high means (intercepts) also have positive slopes (for that variable).
- Can school-level effects explain the variability (variances and covariances)?
Tau Matrix (example: 2x2)

One predictor at level 1; thus, $T$ contains variances and covariances for the (randomly varying) intercepts and slopes

$$T = \begin{bmatrix}
\text{var}(u_{0j}) & \text{cov}(u_{0j}, u_{1j}) \\
\text{cov}(u_{0j}, u_{1j}) & \text{var}(u_{1j})
\end{bmatrix} = \begin{bmatrix}
\tau_{00} & \tau_{01} \\
\tau_{10} & \tau_{11}
\end{bmatrix}$$

Proportion reduction in variance at level 1

- “Variance explained” at level 1 (within schools) by the set of variables
- R & B p. 79

$$pve_{-1} = \frac{\sigma_B^2 - \sigma_F^2}{\sigma_B^2}$$
Intercepts- and Slopes-as Outcomes

• The most general form of multilevel model
• Model both the intercepts and slopes with explanatory variables at level 2
• Trying to account for (reduce) variance in the intercepts and slopes, based on contextual characteristics
• Cross-level interactions- does the effect of a contextual variable moderate the impact of a level 1 variable on the dependent variable?

Using context to predict the intercept AND the slope

Level 1
\[ Y_{ij} = \beta_{0j} + \beta_{1j}(SES)_{ij} + r_{ij} \]

Level 2
\[ \beta_{0j} = \gamma_{00} + \gamma_{01}(Sector)_{j} + u_{0j} \]
\[ \beta_{1j} = \gamma_{10} + \gamma_{11}(Sector)_{j} + u_{1j} \]
Combining terms...

\[ Y_{ij} = \gamma_{00} + \gamma_{10}SES_{ij} + \gamma_{01}Sector_j + \gamma_{11}SES_jSector_j + u_{0j} + u_{1j}SES_{ij} + r_{ij} \]

Proportion reduction in variance at level 2

- “Variance explained” in intercept at level 2 (between schools) by the set of variables
- R & B p. 74

\[ pve_{\beta_0j} = \frac{\tau_{00B} - \tau_{00F}}{\tau_{00B}} \]
Proportion reduction in variance at level 2

• “Variance explained” in slope at level 2 (between schools) by the set of variables
• R & B p. 74

\[
pve_{\beta_{1j}} = \frac{\tau_{11B} - \tau_{11F}}{\tau_{11B}}
\]

Error Terms in Level 2

• \(u_{0j}\) and the \(u_{qj}\) are now the residuals for each estimated level-1 effect (intercepts, slopes) (i.e., the level-2 models).
• The variance for these errors represents the variance remaining in the level-1 intercept and slopes after controlling (accounting) for the W’s.
  • Not independent within schools, because the level-2 errors are common to every student in a particular school
  • These errors vary across schools, and depend on the level-2 IV’s which vary across schools.
Variance Components

- $\text{Var}(u_{0j}) = (\tau_{00})$
- $\text{Var}(u_{qj}) = (\tau_{qq})$
- $\text{Cov}(u_{0j}, u_{qj}) = \tau_{0q}$

• For competing models, reducing the variability of these error terms implies that the level-2 (group or context) variables provides some explanation of the dependent variable.
• Consider choosing the model that has best explanation (reduction) of variance

References—My favorite books