Multilevel Modeling: Applications and Procedures

D. Betsy McCoach
University of Connecticut

Scott J. Peters
Purdue University

Matthew T. McBee
Frank Porter Graham Child Development Institute
University of North Carolina-Chapel Hill

Multilevel Models: Two main types

- Longitudinal Models: observations clustered within people
  - Within-Person Variation Models (change within a person not due to time)
  - Growth Curve Models
- Organizational Models: People clustered within groups
  - Students within classes/schools
  - family/dyad data

Scenario One

An administrator is interested in the relationship between hours of English instruction and student Verbal standardized test score. The administrator is also interested in whether females are affected differently than men. The study is carried out in 300 schools nationwide, with approximately 50-60 students per school. What are the variables involved here? Describe how you would analyze these data to answer the questions of interest.

Question

- Why can't we simply measure everybody together and run some multivariate stats?

Answer: Dependency

- There is dependency among observations, and this needs appropriate modeling. And MLM provides such appropriate modeling.

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.

Answer: 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)

The Nesting Problem

- Observations which are nested within larger units 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 a traditional linear models approach.
- Even mild violations can lead to severe problems with Type I error (typically inflated).
  - Kish (1965): Cluster effect

Scenario Two

In the same 300 schools around the U.S., students are sampled. This sampling yielded 50 to 60 children per school, who were then assessed on 5 occasions throughout the school year on reading achievement. Several other variables were measured: school size, student IQ (WISC-R), and a measure of a child's attention ability which was assessed at each of the 5 occasions. Identify the outcome and the predictors. Describe how you would analyze these data.

Question

- What is the problem with trying to assess students in a school with multiple measures at several time points?
  - Keep in mind that many of these time points might be different for each student and might not include the same information.
  - Think about your GLM/ANOVA/Regression assumptions

Answer

- MLM does not require a balanced design in measurement occasions or variables. MLM allows for time unstructured data.
- Even one time point out of seven can still be used in the model!

Missing Data

- If MCAR or MAR = no problem if in response variable
- If MNAR, there are pattern-mixture models and selection models to account for this (SAS PROC NLMIXED / GLIMIXED / MPlus)
- Imputation is also possible if missing data is in predictor variables (MCAR / MAR)
- Every data point is usable!

Reasons for using HLM…

- Assumes a realistic error structure for clustered data, yielding correct standard errors
- 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

The Nesting Problem

- Think about the regression model and its assumptions

$$\mathbf{y} = \mathbf{BX} + \mathbf{e}$$
$$\mathbf{e} \sim N(0, \sigma^2)$$

- The variance-covariance matrix of **e** is assumed to be $\sigma^2\mathbf{I}$.
- This means an identity matrix (**I**) with σ^2 on the diagonal and zeros everywhere else.

The Nesting Problem

- The covariance matrix of **e**, which is assumed to be $\sigma^2\mathbf{I}$, looks like this:

$$\begin{bmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 & 0 & 0 \; \ldots\ldots \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2 \\ \ldots\ldots\ldots & & & & & & \end{bmatrix}$$

The nesting problem

- The problem with this assumption is the zeros.
- These zeros say that, according to the model, there is no correlation between individual scores after the predictor variables have been accounted for.
- However, we know that in nested data, this isn't true.

Multilevel models and nesting

- Multilevel models make different assumptions about the errors.
- Typically the variance is decomposed into two parts, one representing variance within clusters and another representing variance across clusters.
- Within-cluster variance: $r_{ij} \sim N(0, \sigma^2)$
- Across-cluster variance: $u_{0j} \sim N(0, \tau_{00})$

Multilevel Models and Nesting

- We now have two residuals in our model.
  - These are typically referred to as "variance components" in multilevel modeling.
- The first, r_{ij}, represents variability within clusters.
- The second, u_{0j}, represents variability between clusters.
- The covariance matrix of the residuals correctly accounts for the nested data structure.

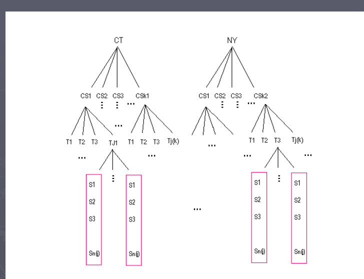
Covariance Matrix of Residuals in Multilevel Models

$$\begin{bmatrix} \sigma^2+\tau_{00} & \tau_{00} & \tau_{00} & 0 & 0 & 0 & \\ \tau_{00} & \sigma^2+\tau_{00} & \tau_{00} & 0 & 0 & 0 & \\ \tau_{00} & \tau_{00} & \sigma^2+\tau_{00} & 0 & 0 & 0 & \\ 0 & 0 & 0 & \sigma^2+\tau_{00} & \tau_{00} & \tau_{00} & \cdots \\ 0 & 0 & 0 & \tau_{00} & \sigma^2+\tau_{00} & \tau_{00} & \\ 0 & 0 & 0 & \tau_{00} & \tau_{00} & \sigma^2+\tau_{00} & \\ \ldots\ldots & & & & & & \end{bmatrix}$$

Multilevel Models and Nesting

- This variance-covariance matrix summarizes our theory about the residuals.
- Within clusters, we assume that individual scores are correlated even after accounting for the predictors.
- We assume that individuals in different clusters are uncorrelated.
- Finally, the total residual variance for each individual is the sum of the within-cluster residual and the between-cluster residual.

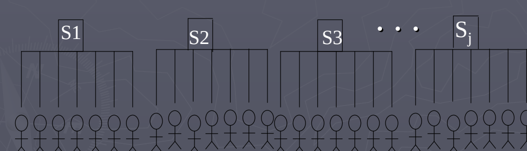
Multi-level Data Structures



Students, nested within teachers, nested within charter schools (nested within States)

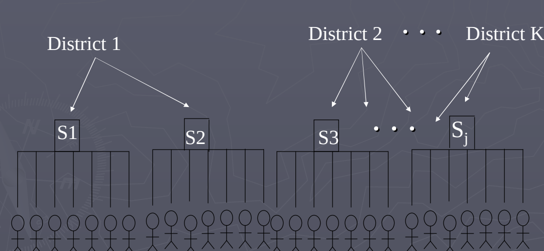
Two-level structure

Children, nested within schools

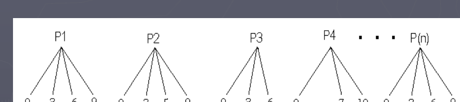


Three-level structure

When groups of schools belong to certain districts:



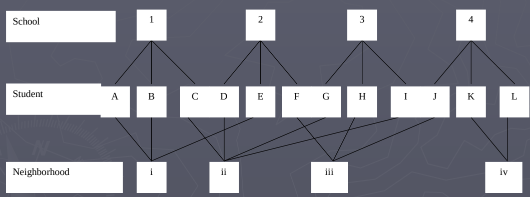
Repeated measures structures



Data collected over months; some participants provide only partial data to the model.

This is an example of "time-unstructured" data, since collection schedules are variable across participants.

Cross-classified structures



Students, cross-classified by school and neighborhood

Beretvas, S. N. (in press)

What does this do for us?

- Allows for quantification of within- and between-level variability (differences)
  - e.g., How much variability in student achievement is explained by school? (Is between schools)? How much of the variability is within school?
  - e.g., How much did a person change within himself / herself v. as compared to others?

It's all just regression…

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

v.

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij}, +u_{0j} + e_{ij}$$

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….

$$VarY_{ij} = Var(u_{0j} + r_{ij}) = \tau_{00} + \sigma^2$$

- The total variability in the dependent variable can be separated into two pieces: that which lies between clusters (τ_{00}) and that which is within clusters σ^2

Intra-class correlation

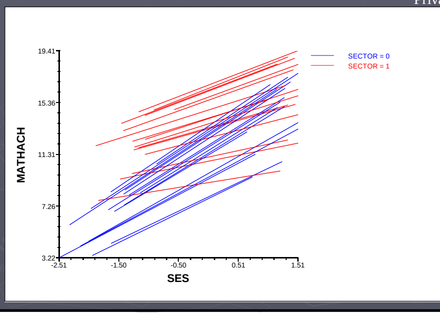
$$ICC = \frac{Var(u_{0j})}{Var(u_{0j} + r_{ij})} = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

- 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)

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

Public are blue (0)

Private are red (1)



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 (raw-score form, or un-centered variable)
  - 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

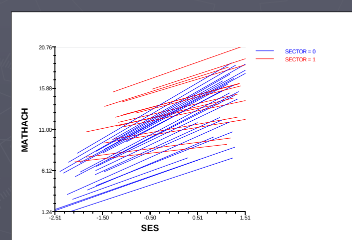
- 2 level-2 equations
  - One equation models the variability/differences in the 160 intercepts
  - One equation models the variability/differences in the 160 SES slopes

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Sector})_{1j} + u_{0j}$$

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

Combining terms…

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



Non-random and random sources of variation for level-1 coefficients

Note: a level-one coefficient may also have both non-random and random sources of variation:

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_{1j} + u_{1j}$$

non-random → "fixed effects"

random

Tau Matrix (example: 2x2)

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

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

Questions?