STATISTICAL HORIZONS

Multilevel Modeling: A Second Course

Kristopher Preacher, Ph.D.

Upcoming Seminar:

February 7 - March 7, 2022, On Demand Seminar

STATISTICAL HORIZONS

What this workshop will accomplish

I will review the basics of multilevel modeling (MLM) to bring everyone up to speed.

I will discuss some more advanced ideas and techniques, and show how to implement them in Mplus.

MSEM will be then explored in depth, with an emphasis on practical application using Mplus.

By the end, you will be able to develop and run your own models, by adapting and combining some of the example syntax provided in the online workshop materials.

5

STATISTICAL HORIZONS

What this workshop isn't

This is not a basic Mplus workshop

- We will cover some Mplus basics, but not everything—only those features that I have found the most useful for implementing MSEM.

This is not an introductory SEM or MLM course

- I will cover basic topics as they pertain directly to multilevel SEM.
- But if you are a complete newcomer to either or both of SEM or MLM, I recommend that you supplement this workshop with books or workshops dedicated expressly to these topics.

6

STATISTICAL HORIZONS

Review of Multilevel Modeling (MLM): Background

STATISTICAL HORIZONS

Key assumptions of (single-level) multiple linear regression (MLR)

- Correct specification of relationship between **IV**s and **DV**
- Inclusion of the important IVs in the model
- Perfect reliability of IVs
- Constant variance of errors* (homoscedasticity)
- Independence of errors*
- Normality of errors*

$$\left.\begin{array}{l}\\\\\\\\\end{array}\right\} e_i \overset{iid}{\sim} N\left(0, \sigma_e^2\right)$$

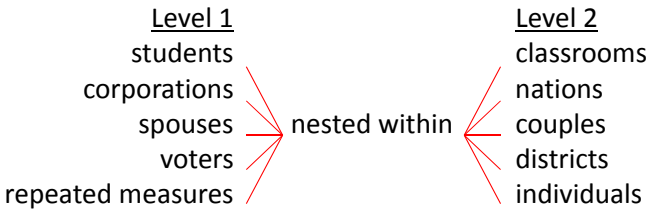
**Errors* are unobserved population quantities. *Residuals* are the corresponding observable sample quantities.

If the relationship between the IVs and DV is correctly specified, then *e* represents random error.

STATISTICAL HORIZONS

Clustered data commonly occur when cases are "nested" in ways that enhance the probability of similar responses.

Examples:



Clustered data violate the assumption of independence, often leading to biased standard errors.

STATISTICAL HORIZONS

Ignoring violations of assumptions caused by nesting means we operate under the assumption that we have more information than we really do.

In clustered data, sampling units no longer yield unique information.

Nesting can be a nuisance from a statistical point of view, but the dependence can also be a source of great substantive interest.

STATISTICAL HORIZONS

Multilevel modeling (MLM) is an elaboration of multiple regression that is designed for use with clustered data.

Also known as hierarchical linear modeling (HLM), random coefficient modeling, contextual analysis, mixed linear modeling, and mixed effects modeling.

Very popular in psychology, education, organizational research, and public health.

11

STATISTICAL HORIZONS

MLM treats clusters as if they are sampled from a larger population of clusters, enhancing the generalizability of results.

In MLM, regression intercepts and/or slopes are assumed to have a particular distribution across clusters, summarized by a limited set of parameters (e.g., mean and variance).

- MLM extends single-level regression by treating these intercepts and slopes as dependent variables in their own right.

In MLM, the dependent variable (y_{ij}) is always measured at the lowest level, Level 1.

12

STATISTICAL HORIZONS

For example, whereas the single-level regression model might assume that all cases come from a population with the same intercept β_0:

$$y_i = \beta_0 + \beta_1 x_i + e_i \qquad e_i \sim N\left(0, \sigma_e^2\right)$$

...MLM permits individual level-2 units to have their own distribution of β_0. Rather than estimate each one individually, we assume a distributional form for β_0 as we do for e_{ij}:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + e_{ij} \qquad e_{ij} \sim N\left(0, \sigma_e^2\right)$$
$$\beta_{0j} \sim N\left(\gamma_{00}, \tau_{00}\right)$$

STATISTICAL HORIZONS

Advantages of MLM

Parsimony. Compared to a fixed effects approach (including $J - 1$ separate level-2 effects of cluster identifiers using dummy codes), we simply estimate a mean and variance for an effect across the population of level-2 units.

Generalizability. Because the effect is treated as random rather than fixed, we can generalize results to the population of level-2 units rather than only to those sampled.

Appropriateness. Multilevel models often conform more closely to theoretical predictions than do other approaches.

Flexibility. Rather than treating clusters as nuisances that violate assumptions, by modeling nestedness we can examine both level-1 and level-2 effects.

STATISTICAL HORIZONS

Review of Multilevel Modeling (MLM): Equations

STATISTICAL HORIZONS

There are two popular ways to represent MLM in equation form:

Matrix expression: The model is expressed for a typical level-2 unit (*cluster*), where level-1 units are assembled into a vector.

Scalar expression: The model is expressed for a typical level-1 unit and a typical level-2 unit.

We will use the scalar expression here.

STATISTICAL HORIZONS

Scalar expression

Level-1 expression...

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + e_{ij}$$

Level-2 expression...

$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_{1j} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_{1j} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}w_{1j} + u_{2j}$$

Combined (reduced) form...

$$y_{ij} = \gamma_{00} + \gamma_{01}w_{1j} + \gamma_{10}x_{1ij} + \gamma_{11}x_{1ij}w_{1j} + \gamma_{20}x_{2ij} + \gamma_{21}x_{2ij}w_{1j} + u_{0j} + u_{1j}x_{1ij} + u_{2j}x_{2ij} + e_{ij}$$

17

STATISTICAL HORIZONS

Scalar expression

Using scalar expressions, the relationship of MLM to multiple linear regression (MLR) becomes clear.

MLM can be thought of as a form of MLR in which the intercept and slopes are potentially random.

We assume these coefficients are normally distributed, and estimate the means, variances, and covariances of their joint distribution as parameters.

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_{1j} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_{1j} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}w_{1j} + u_{2j}$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{20} & \tau_{21} & \tau_{22} \end{pmatrix}\right]$$

$$e_{ij} \sim N\left(0, \sigma^2\right)$$

18

STATISTICAL HORIZONS

The simplest multilevel model: Random effects ANOVA

$$y_{ij} = \beta_{0j} + e_{ij} \quad \text{"Level-1 model"} \qquad e_{ij} \sim N\left(0, \sigma_e^2\right)$$
$$\beta_{0j} = \gamma_{00} + u_{0j} \quad \text{"Level-2 model"} \qquad u_{0j} \sim N\left(0, \tau_{00}\right)$$

The *reduced form equation* for the random effects ANOVA:

$$y_{ij} = \underbrace{\overbrace{\gamma_{00}}^{\text{fixed component}} + \overbrace{u_{0j}}^{\text{random component}}}_{\text{random coefficient } \beta_{0j}} + e_{ij}$$

The between- and within-cluster variances sum to the observed variance:

$$\hat{\tau}_{00} + \hat{\sigma}_e^2 = \hat{\sigma}_y^2$$

β_{0j} is treated as a random coefficient because (a) we are interested not in individual intercepts but rather in the mean (γ_{00}) and variance (τ_{00}) and (b) we want to generalize results beyond the particular groups in our study.

STATISTICAL HORIZONS

The **intraclass correlation (ICC)** is the proportion of variance in *y* that is across level-2 units:

$$\text{ICC} = \frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \hat{\sigma}_e^2}$$

ICC is sometimes used to decide if MLM would be worthwhile.

STATISTICAL HORIZONS

We can expand the random effects ANOVA model to include level-1 predictors. The result is a random effects ANOVA model with covariates (i.e., RANCOVA):

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$

The effect of *x* is constrained to be the same across all level-2 units. In other words, $\tau_{11} = 0$ (where τ_{11} is the slope variance). The intercept variance (τ_{00}), on the other hand, is allowed to be nonzero.

The reduced-form equation is: $y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j} + e_{ij}$

The level-1 variance σ_e^2 is now the residual variance after adjusting for the predictor *x*. The level-2 variance t_{00} is the group-level residual variance.

STATISTICAL HORIZONS

We could expand the RANCOVA model to include more level-1 predictors or level-2 predictors. For example:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \beta_{3j}x_{3ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_{1j} + \gamma_{02}w_{2j} + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$

STATISTICAL HORIZONS

In principle, we can have any number of level-1 predictors, and any mix of fixed and random coefficients. For example:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \beta_{3j}x_{3ij} + \beta_{4j}x_{4ij} + \beta_{5j}x_{5ij} + \beta_{6j}x_{6ij} + \beta_{7j}x_{7ij} + e_{ij}$$

$$\begin{aligned}
\beta_{0j} &= \gamma_{00} + u_{0j} \\
\beta_{1j} &= \gamma_{10} + u_{1j} \\
\beta_{2j} &= \gamma_{20} + u_{2j} \\
\beta_{3j} &= \gamma_{30} + u_{3j} \\
\beta_{4j} &= \gamma_{40} + u_{4j} \\
\beta_{5j} &= \gamma_{50} + u_{5j} \\
\beta_{6j} &= \gamma_{60} + u_{6j} \\
\beta_{7j} &= \gamma_{70}
\end{aligned}$$

$$e_{ij} \sim N\left(0, \sigma_e^2\right)$$

$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \\ u_{3j} \\ u_{4j} \\ u_{5j} \\ u_{6j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & & & & & \\ \tau_{10} & \tau_{11} & & & & & \\ \tau_{20} & \tau_{21} & \tau_{22} & & & & \\ \tau_{30} & \tau_{31} & \tau_{32} & \tau_{33} & & & \\ \tau_{40} & \tau_{41} & \tau_{42} & \tau_{43} & \tau_{44} & & \\ \tau_{50} & \tau_{51} & \tau_{52} & \tau_{53} & \tau_{54} & \tau_{55} & \\ \tau_{60} & \tau_{61} & \tau_{62} & \tau_{63} & \tau_{64} & \tau_{65} & \tau_{66} \end{bmatrix}\right)$$

All these slopes have the same interpretations as in single-level regression.

STATISTICAL HORIZONS

Random effects can exist at any level *except the highest level*. For example, level-1 slopes can vary randomly across level-2 units, but in a two-level model there are no level-3 units for level-2 slopes to vary across.

$$\begin{aligned}
y_{ij} &= \beta_{0j} + \beta_{1j}x_{1ij} + e_{ij} \\
\beta_{0j} &= \gamma_{00} + \gamma_{01}w_{1j} + u_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}w_{1j} + u_{1j}
\end{aligned}$$

STATISTICAL HORIZONS

Model building strategies

There are three common strategies for model-fitting in MLM:

- **A priori specification**. Decide on the model based exclusively on theory.
- **Build-up strategy**. Add predictors and random effects one or a few at a time, and retain if significant and/or meaningfully large.
- **Tear-down strategy**. Start with maximally complex model, and remove predictors and random effects that are not significant or large.

A priori specification is the most scientifically defensible. The build-up strategy is perhaps the most common in practice.

STATISTICAL HORIZONS

For example, using the **build-up strategy**, we might progress through these stages, each time examining the parameter estimates, statistical significance, differences in model fit, and changes in explained variance.

$$y_{ij} = \beta_{0j} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$\beta_{2j} = \gamma_{20}$$

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + u_{2j}$$

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}w_{1j} + \gamma_{02}w_{2j} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}w_{2j} + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}w_{1j} + \gamma_{22}w_{2j} + u_{2j}$$

STATISTICAL HORIZONS

Expanding the number of random coefficients expands size of the tau (**T**) matrix rapidly, increasing the risk of estimation errors and lowering parsimony:

Random intercept only: $u_{0j} \sim N(0, \tau_{00})$

Random intercept and slope: $\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix}\right)$

Random intercept, 2 slopes: $\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix}\right)$

Random intercept, 4 slopes: $\begin{bmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \\ u_{3j} \\ u_{4j} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & & & \\ \tau_{10} & \tau_{11} & & & \\ \tau_{20} & \tau_{21} & \tau_{22} & & \\ \tau_{30} & \tau_{31} & \tau_{32} & \tau_{33} & \\ \tau_{40} & \tau_{41} & \tau_{42} & \tau_{43} & \tau_{44} \end{bmatrix}\right)$

27

STATISTICAL HORIZONS

Centering refers to subtracting a value from raw data. This value is usually the **grand mean**, the **group mean**, or a **reference value** of particular interest.

Reasons why centering is used:

- to facilitate interpretation
- to avoid multicollinearity
- to separate effects into within- and between-cluster components

Centering is widely recommended in a variety of situations.

28

STATISTICAL HORIZONS

Centering: Consequences for parameter estimation

Intercepts are interpreted as the predicted value of *y* where all predictors = 0.

In models *with no* higher-order terms (squared terms, products, etc.), centering changes only the intercept, leaving slopes unaffected.

In models *with* higher-order terms, centering affects all terms *except* the highest-order terms.

STATISTICAL HORIZONS

Grand mean centering: Subtracting the grand mean from all *x* scores.

$$x_{ij} - \overline{x}_{..}$$

$x_{ij} - \overline{x}_{..} = 0$ represents the average level-1 unit's x_{ij} score.

γ_{00} now represents the mean of the predicted *y*s at the grand mean of x_{ij}.

τ_{00} is now the variance of the predicted *y*s at the grand mean of x_{ij}.

STATISTICAL HORIZONS

An example model: Random intercept, random slope

$$y_{ij} = \beta_{0j} + \beta_{1j}\left(x_{ij} - \bar{x}_{..}\right) + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

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

$$y_{ij} = \gamma_{00} + \gamma_{10}\left(x_{ij} - \bar{x}_{..}\right) + u_{1j}\left(x_{ij} - \bar{x}_{..}\right) + u_{0j} + e_{ij}$$

STATISTICAL HORIZONS

Group mean centering: Subtracting the group mean from all *x*s in each group.

$$x_{ij} - \bar{x}_{.j}$$

$x_{ij} - \bar{x}_{.j} = 0$ represents each level-2 unit's average x_{ij} score. Thus, $x_{ij} - \bar{x}_{.j}$ has no level-2 variance.

γ_{00} still represents the mean of the predicted *y*s at the grand mean of x_{ij}.

τ_{00} is the variance of the predicted *y*s at the group means of x_{ij}.

STATISTICAL HORIZONS

An example model: Random intercept, random slope

$$y_{ij} = \beta_{0j} + \beta_{1j}\left(x_{ij} - \bar{x}_{.j}\right) + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$y_{ij} = \gamma_{00} + \gamma_{10}\left(x_{ij} - \bar{x}_{.j}\right) + u_{1j}\left(x_{ij} - \bar{x}_{.j}\right) + u_{0j} + e_{ij}$$

STATISTICAL HORIZONS

Adding the mean back

Group mean centering and using group means as level-2 predictors lets us separate x_{ij}, and thus x_{ij}'s effect, into "within" and "between" components.

$$y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - \bar{x}_{.j}) + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + \gamma_{01}\bar{x}_{.j} + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$
$$y_{ij} = \gamma_{00} + u_{0j} + \underbrace{\gamma_{10}}_{\text{within effect}}\left(x_{ij} - \bar{x}_{.j}\right) + \underbrace{\gamma_{01}}_{\text{between effect}}\bar{x}_{.j} + e_{ij}$$

However, it has been shown that the effect of $\bar{x}_{.j}$ is biased toward that of $(x_{ij} - \bar{x}_{.j})$, especially with small clusters and low ICC (Lüdtke et al., 2008).

STATISTICAL HORIZONS

Centering *y*

In principle, we could split *y* into between and within components as well.

"Between and within" decomposition: A level-1 variable can be thought of as the sum of between-cluster and within-cluster components:

$$y_{ij} = y_{ij} - \bar{y}_{.j} + \bar{y}_{.j}$$

...which seems a bit silly at first, but it is a meaningful split:

$$y_{ij} = \underbrace{\left(y_{ij} - \bar{y}_{.j}\right)}_{\text{varies only within clusters}} + \underbrace{\bar{y}_{.j}}_{\text{varies only between clusters}}$$

STATISTICAL HORIZONS

Centering *y*

$$y_{ij} = \underbrace{\left(y_{ij} - \bar{y}_{.j}\right)}_{\text{varies only within clusters}} + \underbrace{\bar{y}_{.j}}_{\text{varies only between clusters}}$$

Consider individual differences in student achievement within vs. between classrooms. $(y_{ij} - \bar{y}_{.j})$ is a student's achievement relative to his or her peers, whereas $\bar{y}_{.j}$ corresponds to the classroom average achievement.

We will see that it can be interesting to apply different models to $(y_{ij} - \bar{y}_{.j})$ and $\bar{y}_{.j}$.