An Introduction to Multilevel Models

PSYC 943 (930): Fundamentals of Multivariate Modeling
Lecture 25: December 7, 2012
Today’s Class

• Concepts in Longitudinal Modeling

• Between-Person vs. +Within-Person Models

• Repeated Measures ANOVA as MLM

• Introduction to Multilevel Models

• Fixed and Random Effects of Time and Persons

• Families of Models for Change
Dimensions for Organizing Models

• **What kind of outcome variable?**
  – Normal/continuous? → “general linear models (GLM)”
  – Non-normal/categorical? → “generalized linear models”

• **What kind of predictors?** (names relevant only within GLM)
  – Continuous predictors? → “Regression”
  – Categorical predictors? → “ANOVA”

• **How many dimensions of sampling are in your data?**
  – How many ways do your observations differ from each other?
  – What kind of “dependency” or “correlation” is in your data?

Multilevel models are used to quantify and predict dependency in an outcome related to different dimensions of sampling.
What is a Multilevel Model (MLM)?

• **Same as other terms you have heard of:**
  – **General Linear Mixed Model** (if you are from statistics)
    • *Mixed* = Fixed and Random effects
  – **Random Coefficients Model** (also if you are from statistics)
    • Random coefficients = Random effects
  – **Hierarchical Linear Model** (if you are from education)
    • Not the same as hierarchical regression

• **Special cases of MLM:**
  – Random Effects ANOVA or Repeated Measures ANOVA
  – (Latent) Growth Curve Model (where “latent” → SEM)
  – Within-Person Variation Model (e.g., for daily diary data)
  – Clustered/Nested Observations Model (e.g., for kids in schools)
  – Cross-Classified Models (e.g., “value-added” models)
The Two Sides of a Model

• **Model for the Means:**
  – *Aka Fixed Effects*, Structural Part of Model
  – What you are used to *caring about for testing hypotheses*
  – How the expected outcome for a given observation varies as a function of values on predictor variables
  – *Fixed effects are always specified as a function of known things*

• **Model for the Variances (“piles of variance”):**
  – *Aka Random Effects and Residuals*, Stochastic Part of Model
  – What you are used to *making assumptions about* instead
  – Terms that allow model residuals to be related across observations (persons, groups, time, etc) → operate as a function of *sampling*
  – *Random effects and residuals are unknown things ("error")*
  – This is the primary way that MLM differs from the GLM
Data Requirements for Longitudinal Modeling

• Multiple measures from the same person!
  – 2 is minimum, but just 2 can lead to problems:
    • Only 1 kind of change is observable (1 difference)
    • Can’t directly model interindividual differences in change, because no differentiation between real change and measurement error is possible
  – More data is better (with diminishing returns)
    • More occasions \(\rightarrow\) better description of the form of change
    • More people \(\rightarrow\) better estimates of individual differences in change; better prediction of those individual differences
2 Types of Within-Person Variation

- **Within-Person Change**: Systematic change
  - Magnitude or direction of change can be different across people
  - “Growth curve models” → Time is meaningfully sampled

- **Within-Person Fluctuation**: No systematic change
  - Outcome just varies/fluctuates over time (e.g., emotion, stress)
  - Time is just a way to get lots of data per person
Levels of Inference in Longitudinal Multilevel Modeling

• **Between-Person (BP) Relationships:**
  – **Level-2** – “**INTER**-individual Differences” – Time-Invariant
  – All longitudinal studies begin as cross-sectional studies

• **Within-Person (WP) Relationships:**
  – **Level-1** – “**INTRA**-individual Variation” – Time-Varying
  – Only longitudinal studies can provide this extra information

• Longitudinal studies allow examination of both types of relationships simultaneously (and their interactions)
  – Any variable measured over time usually has both of these sources of variation: BP and WP
An Empty Between-Person Model

\[ Y_i = \beta_0 + e_i \]

Filling in values:
\[ 32 = 90 + -58 \]

Model for the Means

Y Error

Variance:
\[ \frac{\sum (y - y_{pred})^2}{N - 1} \]
Between-Person Model: One Continuous Predictor

\[ y_i = \beta_0 + \beta_1 X_i + e_i \]

Empty Model:
\[ 32 = 90 + -58 \]

Ability (X) Model:
\[ 32 = 29 + 2 \times 9 + -15 \]

Y Error Variance:
\[ \Sigma (y - y_{pred})^2 \]
\[ \frac{N - 2}{N - 2} \]
A More General Linear Model for Between-Person Analysis

\[ y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \beta_3 x_i z_i + e_i \]

Model for the Means (Fixed Effects):

- Each person’s expected (predicted) outcome is a function of his/her values on x and z (and here, of their interaction, too)

- Even though the grand mean is no longer the best guess (best model) for each person’s outcome, we still call it “the model for the means” because each person gets a conditional mean as his or her best guess (same predicted Y given same predictor values)
A More General Linear Model for Between-Person Analysis

\[ y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + e_i \]

**Model for the Variance (Residuals and Random Effects):**

- \( e_i \sim \text{NID} (0, \sigma_e^2) \) → In English: \( e_i \) is a random variable with a mean of 0 and some estimated variance, and is normally distributed.
- **ONE** pile of variance in leftover \( Y \) after accounting for predictors.
- We don’t care about the individual \( e \)'s – we care about their variance.
- What makes this model for the variance “Between-Person”?
  - \( e_i \) is measured only once per person (as indicated by the \( i \) subscript).
Adding Within-Person Variance to the Model for the Variances

Full Sample Distribution: 3 People, 5 Occasions each

Mean = 89.55
Std. Dev. = 15.114
N = 1,334
Empty +Within-Person Model

Start off with Mean of Y as “best guess” for any value:

= Grand Mean

= Fixed Intercept

Can make better guess by taking advantage of repeated observations:

= Person Mean

→ Random Intercept
Division of Error in \textbf{Between Person Model for Variances}
Division of Error in Between Person Model for Variances

$e_{ti}$ represents all Y variance
Division of Error in +Within Person Model for Variances

$U_{0i}$ represents Between-Person Y Variance
$e_{ti}$ represents Within-Person Y Variance

$U_{0i}$ also represents dependency (covariance) due to mean differences in Y across persons
Division of Error in +Within Person Model for Variances
Empty +Within-Person Model

Variance of Y → 2 sources:

**Between-Person Variance:**
- Differences from GRAND mean
- INTER-Individual Differences

**Within-Person Variance:**
- Differences from OWN mean
- INTRA-Individual Differences

Now we have 2 piles of variance in Y to predict.
Between-Person vs. +Within-Person Empty Models

- **Empty Between-Person Model (1 time point):**

  \[ y_i = \beta_0 + e_i \]

  - \( \beta_0 \) = fixed intercept = grand mean
  - \( e_i \) = residual deviation from GRAND mean

- **Empty +Within-Person Model (>1 time points):**

  \[ y_{ti} = \beta_0 + U_{0i} + e_{ti} \]

  - \( \beta_0 \) = fixed intercept = grand mean
  - \( U_{0i} \) = random intercept = individual deviation from GRAND mean
  - \( e_{ti} \) = time-specific residual deviation from OWN mean
The Model for the Variances

• All models have at least an “e” → 1 pile of variance
• 2 main reasons to care about what else should go into the model for the variances:
  – Validity of the tests of the predictors depends on having the ‘right’ model for the variances (where ‘right’ means ‘least wrong’)
    • Estimates will usually be ok → come from model for the means
    • Standard errors (and thus p-values) can be compromised
  – The sources of variation that exist in your outcome will dictate what kinds of predictors will be useful
  – For example, in longitudinal data:
    • Between-Person variation needs Between-Person predictors
    • Within-Person variation needs Within-Person predictors
Categorizing Familiar Models by Their Models for the Variances

• **Multiple Regression, Between-Person ANOVA:** 1 PILE
  - \( y_i = (\beta_0 + \beta_1 X_i + \beta_2 Z_i \ldots) + e_i \)
  - \( e_i \) → ONE residual, assumed uncorrelated with equal variance across observations (here, just persons) → “BP variation”

• **Repeated Measures, Within-Person ANOVA:** 2 PILES
  - \( y_{ti} = (\beta_0 + \beta_1 X_i + \beta_2 Z_i \ldots) + U_{0i} + e_{ti} \)
  - \( U_{0i} \) → A random intercept for differences in person means, assumed uncorrelated with equal variance across persons → “BP variation” = \( \tau_{U_0}^2 \)
  - \( e_{ti} \) → A residual that represents remaining time-to-time variation, usually assumed uncorrelated with equal variance across observations (now, persons and time) → “WP variation” = \( \sigma_e^2 \)
The Model for the Variances

- All models have at least an “e” → 1 pile of variance
- 2 main reasons to care about *what else* should go into the *model for the variances*:
  - **Validity of the tests of the predictors** depends on having the “right” model for the variances (where “right” means “least wrong”)
    - Estimates will usually be ok → come from model for the means
    - Standard errors (and thus *p*-values) can be compromised
  - The sources of variation that exist in your outcome will dictate **what kinds of predictors** will be useful
  - For example, in longitudinal data:
    - Between-Person variation needs Between-Person predictors
    - Within-Person variation needs Within-Person predictors
ANOVA for longitudinal data?

- There are 3 possible “kinds” of ANOVAs we could use:
  - Between-Persons/Groups, Univariate RM, and Multivariate RM

- NONE OF THEM ALLOW:
  - Missing occasions (do listwise deletion instead)
  - Time-varying predictors

- Each includes the same model for the means with respect to time:
  - All possible mean differences (so 4 parameters to get to 4 means)
  - “Saturated means model”: $\beta_0 + \beta_1(T_1) + \beta_2(T_2) + \beta_3(T_3)$
  - The *Time* variable must be balanced and discrete in ANOVA!

- Each kind of ANOVA differs by what it says about the correlation in the data from the same person in the model for the variances…
  - i.e., how it “handles dependency” due to persons, or what it says the variance and covariance of the $y_{ti}$ residuals should look like…
Division of Error in Between Person Model for Variances

$e_{ti}$ represents all $Y$ variance
ANOVA for longitudinal data?

1. Between-Groups Regression or ANOVA
   - BP variance only (1 pile of \( e_{ti} \) only)
   \[
   \begin{bmatrix}
   \sigma_e^2 & 0 & 0 & 0 \\
   0 & \sigma_e^2 & 0 & 0 \\
   0 & 0 & \sigma_e^2 & 0 \\
   0 & 0 & 0 & \sigma_e^2
   \end{bmatrix}
   \]
   - Assumes NO RELATIONSHIP WHATSOEVER among observations from the same person (or across persons)
     • Dependency? What dependency?
     • e.g., 4 occasions * 100 persons would be 400 “independent observations”

   - Will usually be VERY WRONG for longitudinal data
     • BP effects tested against wrong df, WP effects tested against wrong df and wrong variance → messed up SEs → messed up \( p \)-values
     • Will also be wrong for clustered data, although perhaps less so (because the correlation among persons from the same group in clustered data is not as strong as the correlation among occasions from the same person in longitudinal data)
Division of Error in +Within Person Model for Variances

$U_{0i}$ represents Between-Person Y Variance
$e_{ti}$ represents Within-Person Y Variance

$U_{0i}$ also represents dependency (covariance) due to mean differences in Y across persons
ANOVA for longitudinal data?

2. (a) Univariate Repeated Measures ANOVA: \( \text{Var}(e_{ti}) + \text{Var}(U_{0i}) \)

- Assumes a **CONSTANT RELATIONSHIP OVER TIME** among observations from the same person: Compound Symmetry

\[
\begin{bmatrix}
\sigma_c^2 + \tau_{u0}^2 & \tau_{u0} & \tau_{u0} & \tau_{u0} \\
\tau_{u0} & \sigma_c^2 + \tau_{u0}^2 & \tau_{u0} & \tau_{u0} \\
\tau_{u0} & \tau_{u0} & \sigma_c^2 + \tau_{u0}^2 & \tau_{u0} \\
\tau_{u0} & \tau_{u0} & \tau_{u0} & \sigma_c^2 + \tau_{u0}^2
\end{bmatrix}
\]

- Observations from the same person are correlated because of constant person mean differences (via \( U_{0i} \))
  \( \rightarrow \) only 1 kind of person dependency

- Will usually be **SOMEWHAT WRONG** for longitudinal data
  
  - If people change at different rates, the variances and covariances of the outcome over time have to change, too
2. (b) Univariate RM ANOVA with sphericity corrections
   - Based on $\varepsilon \rightarrow$ how far off sphericity, (ranges 0-1, 1=spherical)
   - Applies an overall correction for model df based on estimated $\varepsilon$
   - Corrections for sphericity do not really solve the problem

3. Multivariate Repeated Measures ANOVA
   - Assumes nothing: all variances and covariances are estimated separately \( \rightarrow \) “Unstructured”
   - Because it can never be wrong, an unstructured model can be useful for complete longitudinal data with few occasions
   - Becomes hard to estimate very quickly with many occasions
     - Parameters needed = (\#occasions * [\#occasions+1]) / 2
Example Data Individual Observed Trajectories \((N = 101, n = 6)\)
Summary: ANOVA approaches for longitudinal data are “one size fits most”:

- **Saturated Model for the Means** (balanced time required)
  - All possible mean differences
  - Unparsimonious, but best-fitting (is a description, not a model)

- **3 kinds of Models for the Variances** (complete data required)
  - \( e_{ti} \) only = Between-Person/Group ANOVA → assumes independent data
  - \( U_{0i} \) and \( e_{ti} \) = Compound Symmetry (CS) = Univ. RM ANOVA
    - Requires sphericity, which rarely holds in longitudinal data
  - All possible var. and covar. = Unstructured (UN) = Multiv. RM ANOVA
    - Unparsimonious; is a description, not a model

- **MLM will give us more flexibility in both parts of the model:**
  - Fixed effects that *predict* the pattern of means
  - Random intercepts and slopes and/or alternative covariance structures that *predict* the pattern of variation and covariation over time
Empty Longitudinal Multilevel Model: Review of Terminology

Variance of Y $\rightarrow$ 2 sources:

**Level 2 Random Intercept Variance** (of $U_{0i}$):
- $\rightarrow$ Between-Person Variance ($\tau_{U0}^2$)
- $\rightarrow$ Difference from GRAND mean
- $\rightarrow$ INTER-Individual Differences

**Level 1 Residual Variance** (of $e_{ti}$):
- $\rightarrow$ Within-Person Variance ($\sigma_e^2$)
- $\rightarrow$ Difference from OWN mean
- $\rightarrow$ INTRA-Individual Differences
Empty* Multilevel Model

Model for the Means; Model for the Variance

General Linear Model

\[ y_i = \beta_0 + e_i \]

Multilevel Model

Level 1: \[ y_{ti} = \beta_{0i} + e_{ti} \]

Level 2: \[ \beta_{0i} = \gamma_{00} + U_{0i} \]

3 Model Parameters

1 Fixed Effect:
\[ \gamma_{00} \rightarrow \text{fixed intercept} \]

1 Random Effect (intercept):
\[ U_{0i} \rightarrow \text{person-specific deviation} \]
\[ \rightarrow \text{mean}=0, \text{variance} = \tau_{U0}^2 \]

1 Residual Error:
\[ e_{ti} \rightarrow \text{time-specific deviation} \]
\[ \rightarrow \text{mean}=0, \text{variance} = \sigma_e^2 \]

Composite equation: \[ y_{ti} = \gamma_{00} + U_{0i} + e_{ti} \]

* To be more clear, I call this an “empty means, random intercept” model
Empty Multilevel Model: Useful Descriptive Statistic $\rightarrow$ ICC

Intraclass Correlation (ICC):

$$\text{ICC} = \frac{\text{Intercept Variance}}{\text{Intercept Variance} + \text{Residual Variance}} = \frac{T_{U_0}^2}{T_{U_0}^2 + \sigma_e^2}$$

$$\text{ICC} = \frac{\text{Between-Person Variance}}{\text{Between-Person Variance} + \text{Within-Person Variance}}$$

- ICC = Proportion of total variance that is between persons
- ICC = Average correlation among occasions
- ICC is a standardized way of expressing how much we need to worry about dependency due to person mean differences (i.e., ICC is an effect size for constant person dependency)
Counter-Intuitive: Between-Person Variance is in the numerator, but the ICC is the correlation over time!

\[
\text{ICC} = \frac{\text{Between-Person Variance}}{\text{Between-Person Variance} + \text{Within-Person Variance}}
\]

\[\Rightarrow \text{Large ICC} \Rightarrow \text{Large correlation over time}\]

\[\text{ICC} = \frac{\text{btw}}{\text{btw} + \text{WITHIN}}\]

\[\Rightarrow \text{Small ICC} \Rightarrow \text{Small correlation over time}\]
More New Vocabulary: Labels for 3 Types of Models

- **“Empty Model”** → **“Empty means, random intercept model”**
  - Just fixed intercept, random intercept variance, residual variance
  - First baseline model for everything to follow
  - Used to compute an ICC

- **“Unconditional (Growth) Model”** is up next
  - Effects related to time, but no other predictors yet
  - Second baseline model for everything to follow

- **“Conditional (Growth) Model”** is coming next semester
  - With predictors besides time
  - May end up with more than 1 depending on research questions
Extending the Multilevel Model: 2 Questions for Effects of Time

1. Is there an effect of time **on average**?
   - Is the average line not flat?
   - Significant **Fixed** Effect of Time

2. Does the average effect of time **vary across individuals**?
   - Does each person need their own line?
   - Significant **Random** Effect of Time
Fixed and Random Effects of Time

- No Fixed, No Random
- Yes Fixed, No Random
- No Fixed, Yes Random
- Yes Fixed, Yes Random
Random Linear Time Model

Multilevel Model

Level 1: \[ y_{ti} = \beta_{0i} + \beta_{1i} \text{Time}_{ti} + e_{ti} \]

Level 2: \[ \beta_{0i} = \gamma_{00} + U_{0i} \] \[ \beta_{1i} = \gamma_{10} + U_{1i} \]

Composite Model

\[ y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{1i}) \text{Time}_{ti} + e_{ti} \]
Fixed & Random Effects of Time

\[ y_{ti} = (Y_{00} + U_{0i}) + (Y_{10} + U_{1i}) \text{Time}_{ti} + \epsilon_{ti} \]

**Fixed**
- Intercept

**Random**
- Intercept Deviation
- Slope Deviation
- Int-Slope Covariance

**Model Parameters:**
- **2 Fixed Effects:**
  - \( Y_{00} \) Intercept, \( Y_{10} \) Slope
- **2(+) Random Effects:**
  - \( U_{0i} \) Intercept Variance
  - \( U_{1i} \) Slope Variance
  - Int-Slope Covariance
- **1 Residual Variance**
Unbalanced Time $\rightarrow$ Different time measurements across persons? OK

Rounding time can lead to incorrect estimates!

Code time as **exactly** as possible.

MLM uses each complete time point for each person.

This red predicted slope will probably be made steeper because it is based on less data, though – the term for this is “shrinkage”.

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
Longitudinal Data: Modeling Means and Variances

• We have two tasks in describing the effects of “time”:

1. Choose a Model for the Means
   • What kind of change in the outcome do we have on average?
   • What kind of and how many parameters do we need to represent that change as parsimoniously but accurately as possible?

2. Choose a Model for the Variances
   • What kind of pattern do the variances and covariances of the outcome show over time due to individual differences in change?
   • What kind of and how many parameters do we need to represent that pattern as parsimoniously but accurately as possible?
Big Picture Modeling Framework: Choices for Modeling Means

- What kind of change on average (in the means)?
  - **Empty** refers to model for the means with no predictors (just fixed intercept for grand mean outcome over time)
  - **Saturated** refers to model for means with all possible means estimated (#parameters = #occasions) → THIS IS ANOVA

  - Is a DESCRIPTION of the means, not a predictive MODEL

Even if time is unbalanced!
Choices for Modeling Variance

- The partitioning of variance into piles...
  - Level 2 = BP → **G** matrix of random effects variances/covariances
  - Level 1 = WP → **R** matrix of residual variances/covariances
  - **G** and **R** combine to create **V** matrix of total variances/covariances
  - Many flexible options that allows the variances and covariances to vary in a time-dependent way that better matches the actual data

<table>
<thead>
<tr>
<th>Compound Symmetry (CS)</th>
<th>Unstructured (UN)</th>
</tr>
</thead>
</table>
| \[
\begin{bmatrix}
\sigma_c^2 + \tau_{u_0}^2 & \rho \tau_{u_0}^2 & \rho \tau_{u_0}^2 & \rho \tau_{u_0}^2 \\
\tau_{u_0}^2 & \sigma_c^2 + \tau_{u_0}^2 & \rho \tau_{u_0}^2 & \rho \tau_{u_0}^2 \\
\tau_{u_0}^2 & \tau_{u_0}^2 & \sigma_c^2 + \tau_{u_0}^2 & \rho \tau_{u_0}^2 \\
\tau_{u_0}^2 & \tau_{u_0}^2 & \tau_{u_0}^2 & \sigma_c^2 + \tau_{u_0}^2 \\
\end{bmatrix}
\] | \[
\begin{bmatrix}
\sigma_{11}^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\
\sigma_{21} & \sigma_{22}^2 & \sigma_{23} & \sigma_{24} \\
\sigma_{31} & \sigma_{32} & \sigma_{33}^2 & \sigma_{34} \\
\sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44}^2 \\
\end{bmatrix}
\] |

MLM uses **G** and **R** to predict something in-between!

*Even if time is unbalanced!*
The Point of MLM: Dependency

• Common description of the purpose of MLM is that it ‘addresses’ or ‘handles’ correlated (dependent) data…
• But where does this ‘correlation’ come from? 3 places (here, an example with health as an outcome):

1. Mean differences across persons
   • Some people are just healthier than others (at every time point)

2. Differences in effects of predictors across persons
   • Does time affect health more in some persons than others?
   • Does daily stress affect health more in some persons than others?

3. Non-constant within-person correlation for unknown reasons
   • Occasions closer together may just be more related
   • More likely for outcomes that fluctuate (than change) over time
2-Level Models for the Variances

- Where does the correlation or ‘dependency’ go? Into a new random effects variance component (‘pile of variance’)

- \[ \sigma_e^2 \]
- \[ \tau_{U_0}^2 \]
- \[ \tau_{U_i}^2 \]

- Original Residual Variance \( \sigma_e^2 \)
- Level-1 Residual Variance \( \sigma_e^2 \)
- Level-2 Random Intercept Variance \( \tau_{U_0}^2 \)
- Level-2 Random Linear Slope Variance \( \tau_{U_i}^2 \)

- Within-Person Differences
- Between-Person Differences

\[ \tau_{U_{0i}} \text{ covariance} \]
Summary: The Point of MLM

• All we’ve done so far is carve up our total variance into up to 3 piles:
  – BP (error) variance around intercept
  – BP (error) variance around slope
  – WP (error) residual variance

But making piles does not make error variance go away…

Level 1 (one source of) Within-Person Variation:
Gets accounted for by time-level predictors

Residual Variance \( \sigma_e^2 \)

Level 2 (two sources of) Between-Person Variation:
Gets accounted for by person-level predictors

BP Int Variance \( \tau_{U0}^2 \)

BP Slope Variance \( \tau_{U1}^2 \)

\( \tau_{U01} \) covariance

FIXED effects make variance go away (explain variance).

RANDOM effects just make a new pile of variance.
What can MLM do for you?

1. **Model dependency across observations**
   - Longitudinal, clustered, and/or cross-classified data? No problem!
   - Tailor your model of correlation over time and person to your data

2. **Include categorical or continuous predictors at any level**
   - Time-varying, person-level, group-level predictors for each variance
   - Explore reasons for dependency, don’t just control for dependency

3. **Does not require same data structure for each person**
   - Unbalanced or missing data? No problem!

4. **You already know how (even if you don’t know it yet)!**
   - SPSS Mixed, SAS Mixed, Stata, Mplus, R, HLM…
   - What’s an intercept? What’s a slope? What’s a pile of variance?