Lecture 1:
Review of General Linear Models

- **What is Multilevel Modeling?**
- Between-Person vs. +Within-Person Models
- SPSS, SAS, and STATA GLM Examples
- Summary of ANOVA Options for Longitudinal Data

Dimensions for Organizing Models

- What kind of outcome variable?
  - Normal/continuous? → "general linear models (GLM)"
  - Non-normal/categorical? → "genera*lized* 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 *Any* 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 known predictor variables

- **Model for the Variances:**
  - *Aka* **Random Effects and Residuals**, Stochastic Part of Model
  - What you are used to **making assumptions about** instead
  - How model residuals are distributed and related across observations (across persons, groups, time, etc.)
  - ***This is the primary way that MLM differs from the GLM***

Data Requirements for Multilevel Modeling

- A useful outcome variable:
  - Has an <u>interval scale</u>*
    - → A one-unit difference means the same thing across all points of the scale
    - → Each contributing item has an equivalent scale
  - Has scores with the <u>same meaning</u> over observations
    - → Includes the meaning of construct
    - → Includes how items relate to a scale
    - → Implies measurement invariance

<u>FANCY MODELS CANNOT SAVE BAD VARIABLES.</u>

** For the purposes of these **general** linear mixed models specifically*

How We Will Learn MLM

- "Levels" are defined by the context of a study
  - **Level ≈ a dimension of sampling** (can be nested or crossed)
- We will start with MLM for longitudinal data…
  - Level 1 = variation over time, Level 2 = variation over persons
  - More complex case because of the time dimension
- …We will follow with MLM for clustered data…
  - Level 1 = variation over persons, Level 2 = variation over groups
- … and conclude with MLM for clustered longitudinal data
  - Time (Level 1) within persons (Level 2) within groups (Level 3)

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, person, and group 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 (or you will by Friday)!**
   - Use SPSS Mixed, SAS Mixed, Stata, Mplus, R, HLM, MlwiN…
   - What's an intercept? What's a slope? What's a pile of variance?

3. Does not require same data structure per person (by accident or by design)

ANOVA: uses **multivariate** (wide) data structure:

ID	Sex	T1	T2	T3	T4
100	0	5	6	8	12
101	1	4	7	.	11

People missing any data are excluded (data from ID 101 are not included at all)

MLM: uses **stacked** (long) data structure:

Only rows missing data are excluded

100 uses 4 cases
101 uses 3 cases

ID	Sex	Time	Y
100	0	1	5
100	0	2	6
100	0	3	8
100	0	4	12
101	1	1	4
101	1	2	7
101	1	3	.
101	1	4	11

Time can also be **unbalanced** across people such that each person can have his or her own measurement schedule: Time "0.9" "1.4" "3.5" "4.2"…

Data Requirements for Longitudinal Multilevel Modeling

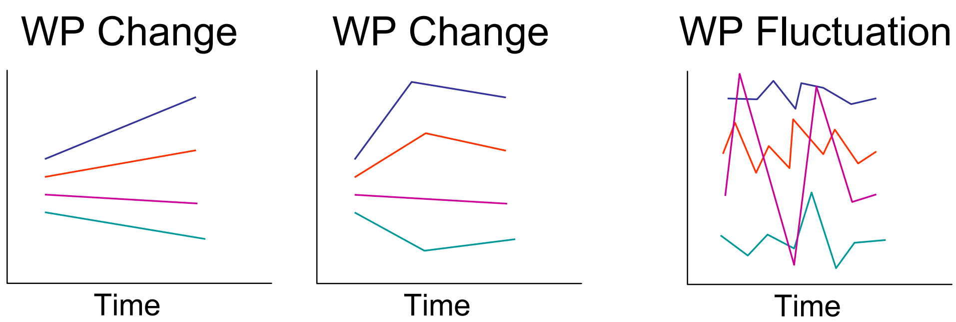
- Multiple measures from the same person!
  - 2 is the minimum, but just 2 can lead to problems:
    - Only 1 kind of change is observable (1 difference)
    - Can't directly model interindividual differences in change, because no differentiation between real change and measurement error is possible
  - More data is better (with diminishing returns)
    - More occasions → better description of the form of change
    - More persons → better estimates of amount of individual differences in change; better prediction of individual differences

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

2 Types of Within-Person Variation

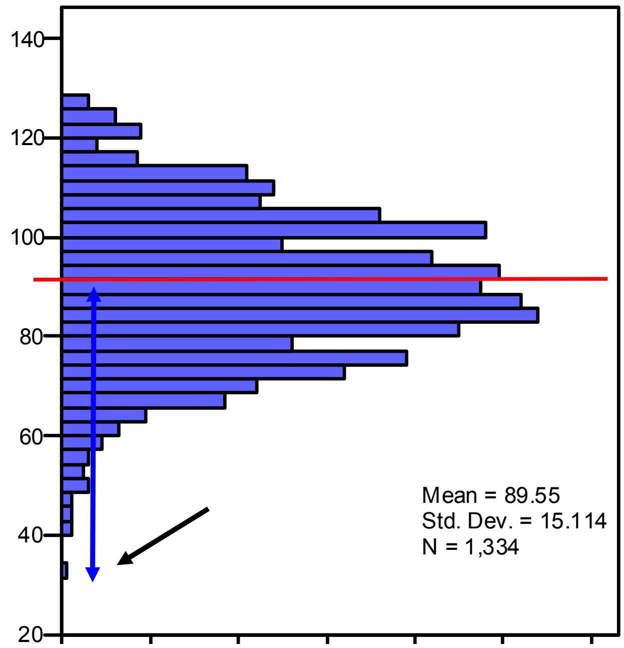
- **Within-Person Change**: Systematic change
  - Magnitude or direction of change can be different across people
  - "Growth curve models" → Time is meaningfully sampled
- **Within-Person Fluctuation**: No systematic change
  - Outcome just varies/fluctuates over time (e.g., emotion, stress)
  - Time is just a way to get lots of data per person



Lecture 1: Review of General Linear Models

- What is Multilevel Modeling?
- **Between-Person vs. +Within-Person Models**
- SPSS, SAS, and STATA GLM Examples
- Summary of ANOVA Options for Longitudinal Data

An “Empty” Between-Person Model



$$Y_i = \beta_0 + e_i$$

Filling in values:

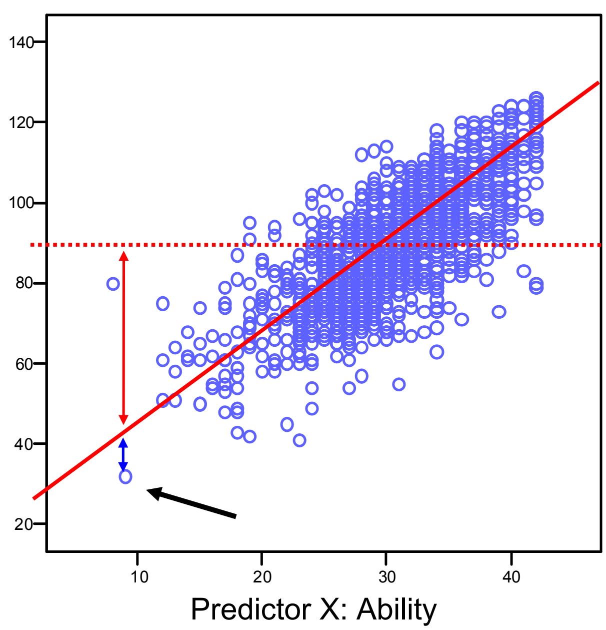
$$32 = \underbrace{90}_{\text{Y pred}} + -58$$

Model for the Means

Y Error Variance:

$$\frac{\Sigma (y - y_{pred})^2}{N - 1}$$

Between-Person Model: One Continuous Predictor



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

Empty Model:

$$32 = 90 + \quad -58$$

Ability (X) Model:

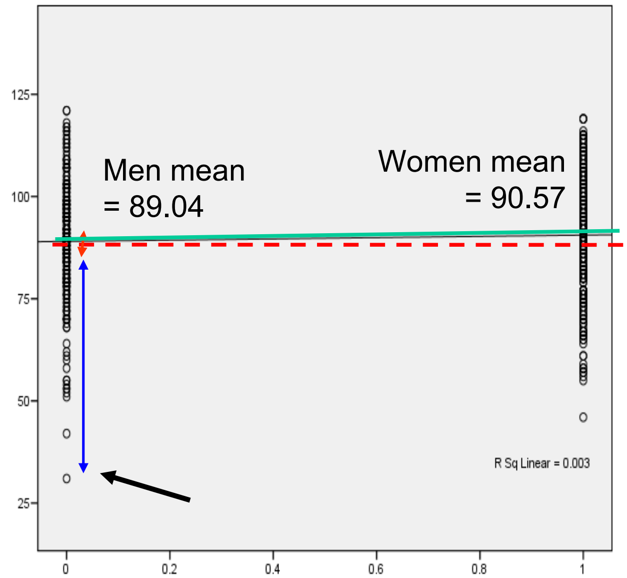
$$32 = \underbrace{29 + 2{*}9}_{\text{Y pred}} + -15$$

Model for the Means

Y Error Variance:

$$\frac{\Sigma (y - y_{pred})^2}{N - 2}$$

Between-Person Model: One Categorical (Grouping) Predictor



Predictor Z: Sex (0=M, 1=W)

$y_i = \beta_0 + \beta_2 Z_i + e_i$

Empty Model:
$32 = 90 + \quad -58$

Sex (Z) Model:
$32 = \underbrace{89 + 1.6{*}0}_{\text{Y pred}} + -57$

Model for the Means

Y Error Variance:

$$\frac{\Sigma\,(y - y_{pred})^2}{N - 2}$$

A General Linear Model for Between-Person Analysis

$$y_i = \boxed{\beta_0 + \beta_1 X_i + \beta_2 Z_i} + e_i$$

Model for the Means (Fixed Effects):

- Each person's expected (predicted) outcome is a function of his/her values on x and z (and potentially 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 the same predictor values)

A General Linear Model for Between-Person Analysis

$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \boxed{e_i}$$

Model for the Variance (Residuals and Random Effects):

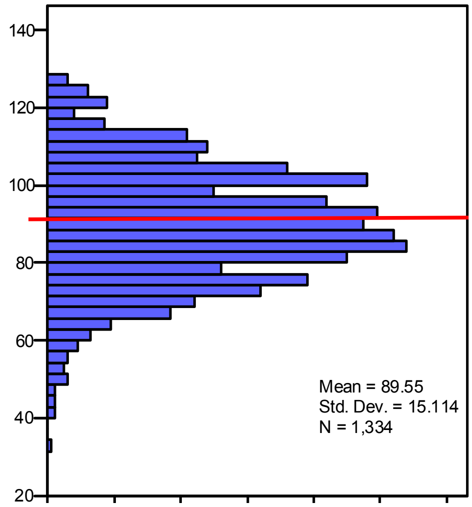
- $e_i \sim NID(0, \sigma_e^2)$ → In English: e_i is a random variable with a mean of 0 and some estimated variance σ_e^2, and is normally distributed
  - e_i is assumed uncorrelated across persons (no covariance)
- **ONE** pile of leftover variance in Y after accounting for predictors
- We don't care about the e values → 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)

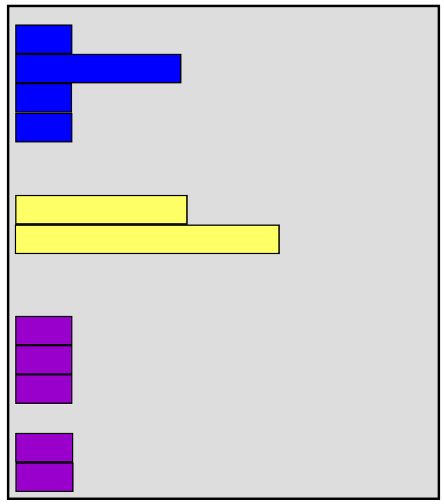
The Two Sides of Each Model

	Model for the Means	Model for the Variances
"Empty"	β_0 → grand mean	e → diff between actual and mean Y
"Regression"	β_0 → Y when X=0 β_1 → ΔY for a 1-unit ΔX	e → diff between actual Y and Y predicted from X
"ANOVA"	β_0 → Y when Z=0 β_2 → ΔY for a 1-unit ΔZ	e → diff between actual Y and Y predicted from Z
"ANCOVA"	β_0 → Y when X=0 & Z=0 β_1 → ΔY for a 1-unit ΔX (constant Z) β_2 → ΔY for a 1-unit ΔZ (constant X)	e → diff between actual Y and Y predicted from X & Z

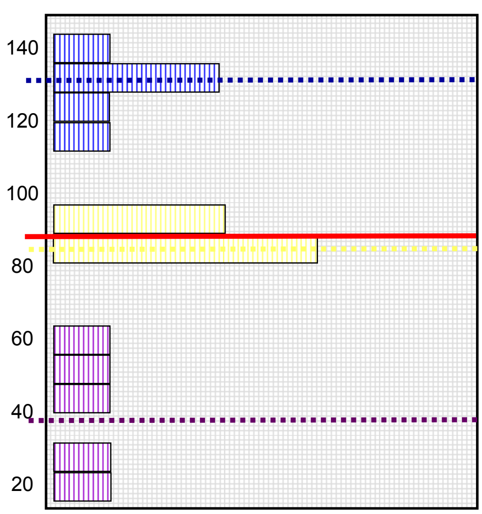
Adding Within-Person Variance to the Model for the Variances

Full Sample Distribution: 3 People, 5 Occasions each





Empty +Within-Person Model



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

= Grand Mean

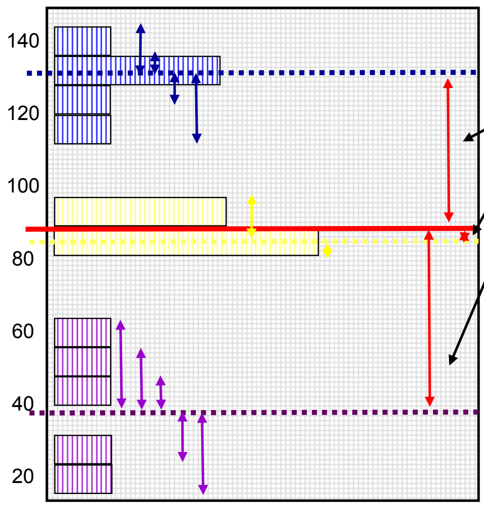
= Fixed Intercept

Can make better guess by taking advantage of repeated observations:

= Person Mean

→ Random Intercept

Empty +Within-Person Model



Variance of Y → 2 sources:

Between-Person Variance:

→ Differences from GRAND mean

→ INTER-Individual Variance:

Within-Person Variance:

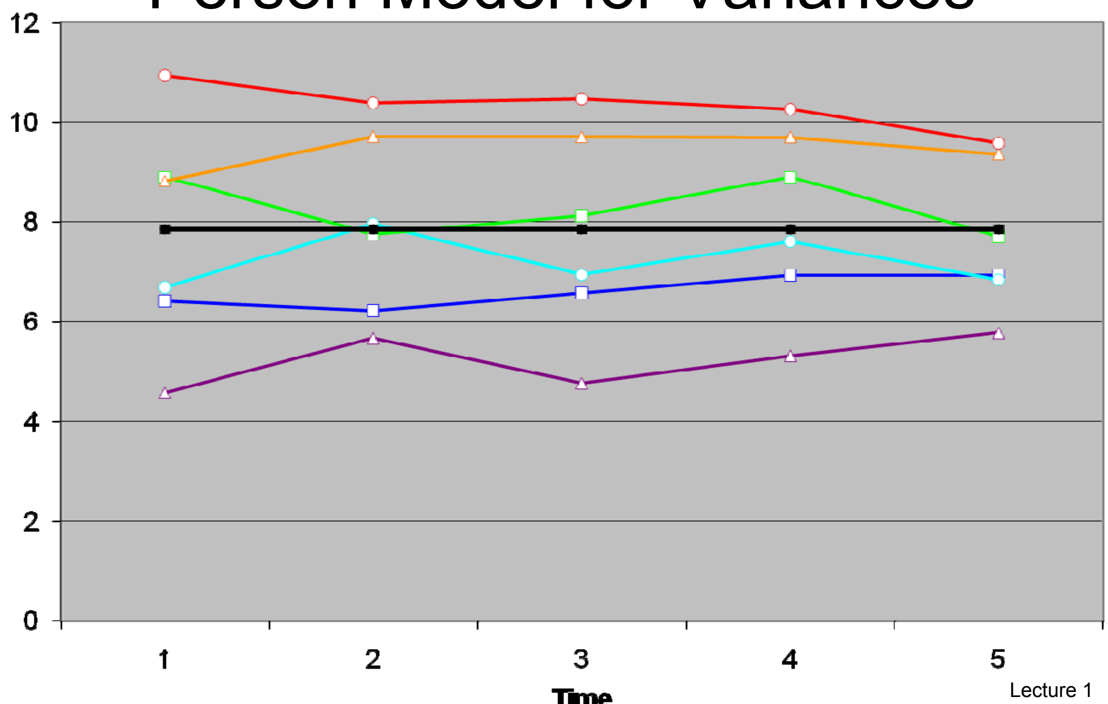
→ Differences from OWN mean

→ INTRA-Individual Differences

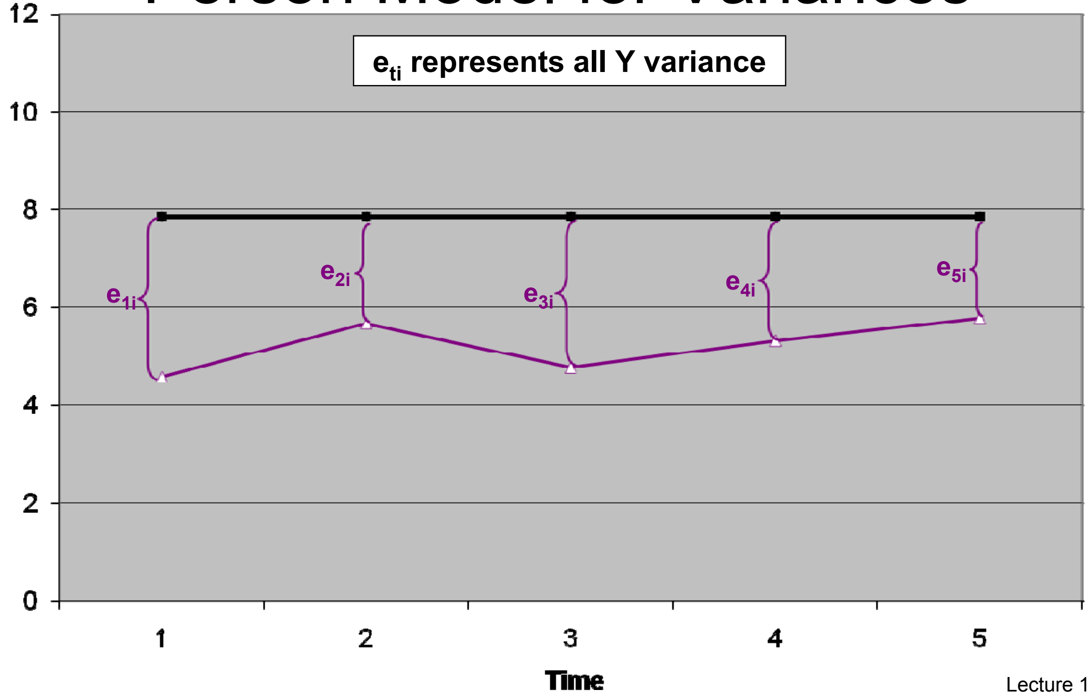
→ *This part is only observable through longitudinal data*

Now we have 2 piles of variance in Y to predict.

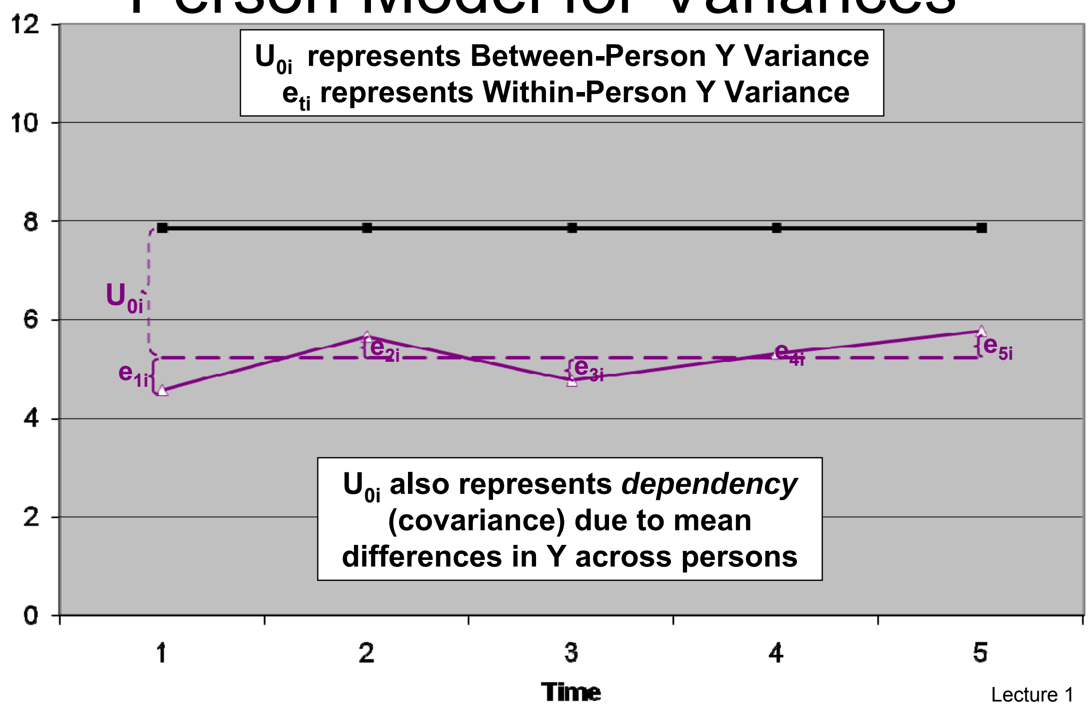
Division of Error in **Between** Person Model for Variances



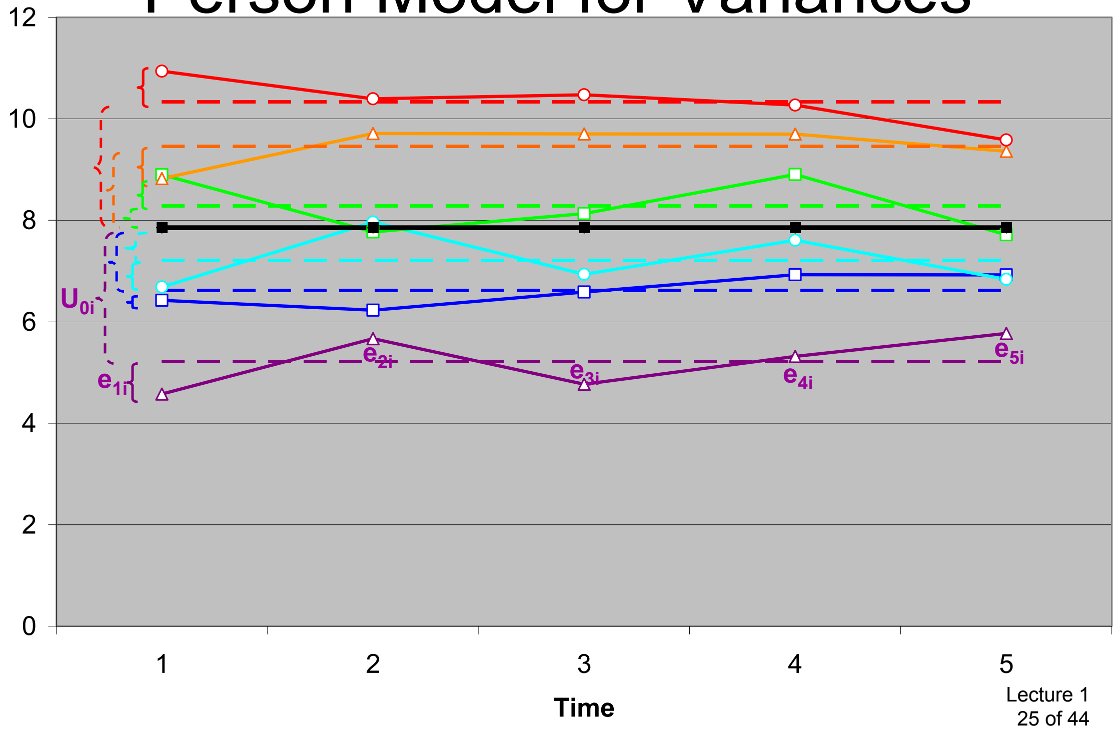
Division of Error in **Between** Person Model for Variances



Division of Error in **+Within** Person Model for Variances



Division of Error in **+Within** Person Model for Variances



Between-Person vs. +Within-Person Empty Models

- <u>Empty Between-Person Model (1 occasion):</u>

$$y_i = \beta_0 + e_i$$

  - β_0 = fixed intercept = grand mean
  - e_i = individual deviation from GRAND mean

- <u>Empty +Within-Person Model (>1 occasion):</u>

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

  - β_0 = fixed intercept = grand mean
  - U_{0i} = random intercept = individual deviation from GRAND mean
  - e_{ti} = individual- and time-specific deviation from OWN mean

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**" = $\tau^2_{U_0}$
  - 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**" = σ^2_e

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

Lecture 1: Review of General Linear Models

- What is Multilevel Modeling?
- Between-Person vs. +Within-Person Models
- **SPSS, SAS, and STATA GLM Examples**
- Summary of ANOVA Options for Longitudinal Data

Statistical models are logically separate from the software that estimates them.

- In this example, we will compare two different models for the variances using different software procedures:
  - **Between-Person Model for Variances**
    - = *Between-Groups* or *Between-Subjects*, or *Independent* ANOVA if all predictors are categorical
    - = *ANCOVA* or *Regression* if some predictors are continuous
    - Estimated via Univariate GLM, Regression, and MIXED/XTMIXED
  - **+Within-Person Model for Variances**
    - = *Within-Subjects*, *Repeated Measures*, or *Dependent* ANOVA if all predictors are categorical and measured within-subjects
    - = *Split-Plot* or *Mixed Design* ANOVA if some categorical predictors are measured between-subjects
    - = *ANCOVA* if some between-person predictors are continuous
    - Estimated via Repeated Measures GLM and MIXED/XTMIXED

Example Data for BP and WP Models

- 100 people ages 10 or 11 (*n* = 50 each)
- Control group vs. treatment group (*n* = 50 each)
- Hypothesis: Outcome should be higher with age, with a greater age difference in the treatment group

Means *(SD)*	Age 10		Age 11	Marginal
Control	49.0 *(4.3)*	5.7	54.7 *(4.0)*	51.9 *(5.0)*
	1.6	**2.8**	4.4	**2.9**
Treatment	50.6 *(3.4)*	8.5	59.1 *(4.3)*	54.8 *(5.8)*
Marginal	49.8 *(3.9)*	**7.1**	56.9 *(4.7)*	**53.3** *(5.6)*

Quick Review: Concepts in building the Model for the Variance

N = # people
t = time
i = person
k = # fixed effects
$\hat{y}_{ti}$ = y predicted from fixed effects

Variance:
Dispersion of y

$$\text{Variance}(y_t) = \frac{\sum_{i=1}^{N} (y_{ti} - \hat{y}_{ti})^2}{N - k}$$

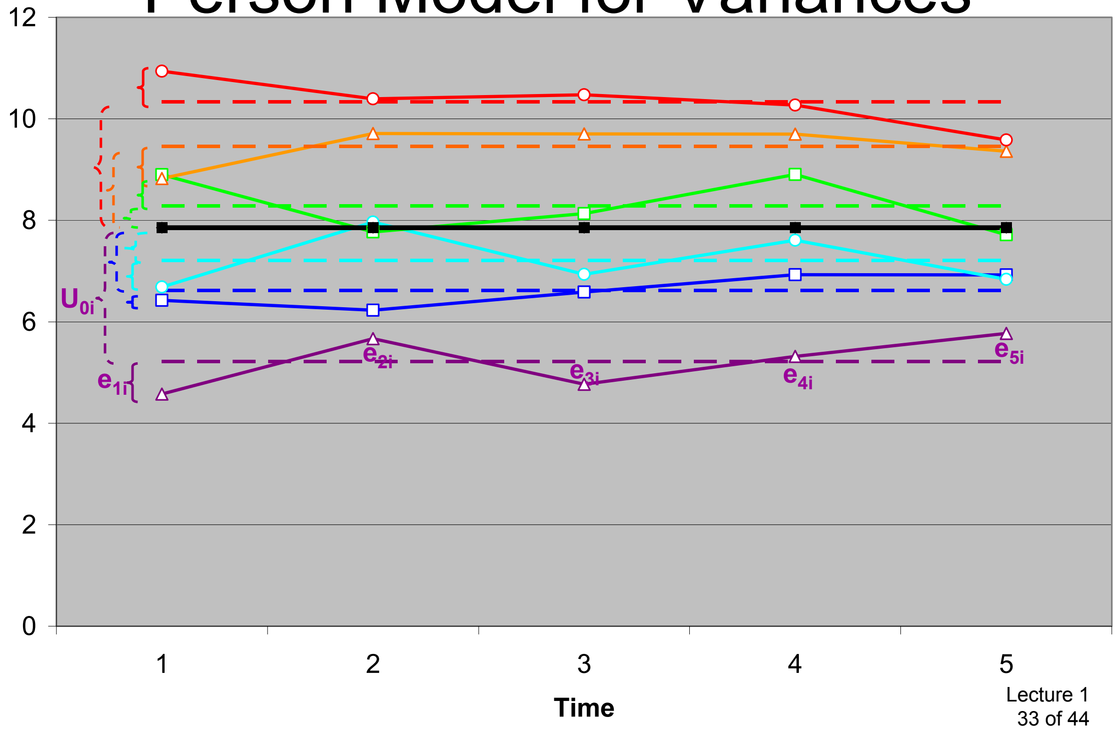
Covariance:
How y values across occasions go together, unstandardized

$$\text{Covariance}(y_1, y_2) = \frac{\sum_{i=1}^{N} (y_{1i} - \hat{y}_{1i})(y_{2i} - \hat{y}_{2i})}{N - k}$$

Correlation:
How y values across occasions go together, standardized (-1 to 1)

$$\text{Correlation}(y_1, y_2) = \frac{\text{Covariance}(y_1, y_2)}{\sqrt{\text{Variance}(y_1)} * \sqrt{\text{Variance}(y_2)}}$$

Division of Error in **+Within** Person Model for Variances



Repeated Measures (+WP) ANOVA: A 2-Pile Model for the Variances

- BP mean differences (U_{0i}) & WP deviations (e_{ti})
- In other words, the variances & covariances of outcomes over time have a specific form → Compound Symmetry
  CS → all variances equal (here, 2), all covariances equal (here, 1)

Data CS Model CS Model in Piles

$$\begin{bmatrix} \sigma_{11}^2 & \sigma_{21} \\ \sigma_{21} & \sigma_{22}^2 \end{bmatrix} \rightarrow \begin{bmatrix} \sigma^2 & \sigma \\ \sigma & \sigma^2 \end{bmatrix} \longrightarrow \begin{bmatrix} \sigma_e^2 + \tau_{u_0}^2 & \tau_{u_0}^2 \\ \tau_{u_0}^2 & \sigma_e^2 + \tau_{u_0}^2 \end{bmatrix}$$

- Total variance at each occasion = Var(e_{ti}) + Var(U_{0i})
- Covariance between occasions = Var(U_{0i})
  - This implies the ONLY reason Time1 and Time 2 are related is because of overall differences between person means

Back to the Example

Data	CS Model	CS Model in Piles
$\begin{bmatrix} \sigma_{11}^2 & \sigma_{21} \\ \sigma_{21} & \sigma_{22}^2 \end{bmatrix}$ →	$\begin{bmatrix} \sigma^2 & \sigma \\ \sigma & \sigma^2 \end{bmatrix}$ →	$\begin{bmatrix} \sigma_e^2+\tau_{u_0}^2 & \tau_{u_0}^2 \\ \tau_{u_0}^2 & \sigma_e^2+\tau_{u_0}^2 \end{bmatrix}$

2-Occasion Example

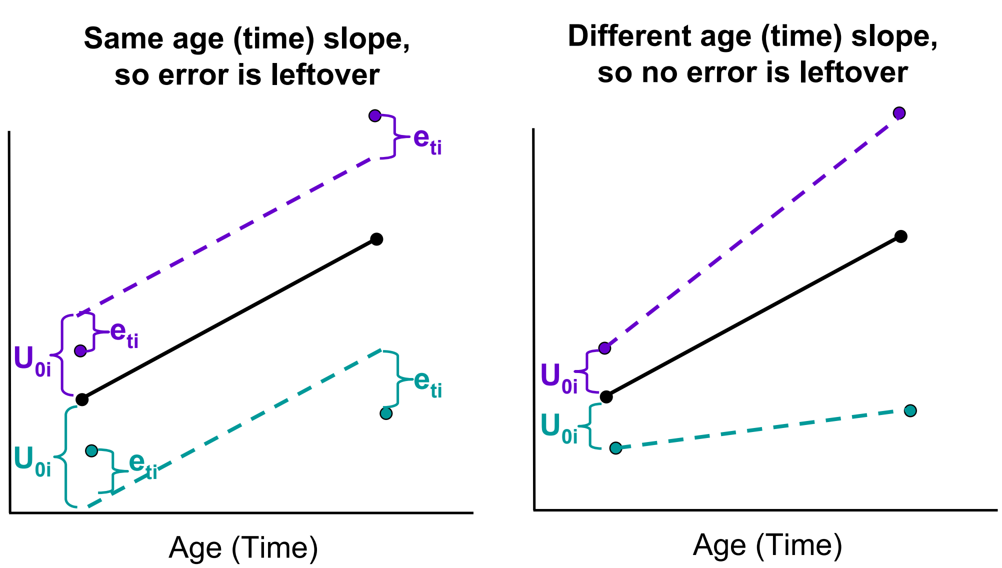
16.08	9.08
9.08	16.08

First, take out fixed effects of age, group, and interaction

Next, decompose residual variance

- Covariance between times (assumed constant; only 1 here) = $Var(U_{0i})$ = 9.08
- Total variance per occasion = $Var(e_{ti}) + Var(U_{0i})$ = 7.00 + 9.08
  - The $Var(U_{0i})$ term (here, 9.08) is what gets kicked out of the denominator for testing any within-person effects (normally doesn't show up in output)
  - The BP effects actually get tested against a larger error term (25.16) than before
- Where does 25.16 (the larger error term used for BP) come from?
  - MSE_{group} = 2N(group mean – grand mean)^2 / df(1)
  - $MSE_{BPerror}$ = 2 (Σ (person mean – group mean)^2) / df(48) = 25.16
    - You are off your group's mean because of YOU (twice) and because of ERROR (once)
  - $MSE_{BPerror}$ = $Var(e_{ti}) + 2*Var(U_{0i})$ → 25.16 = 7.00 + 2*9.08
  - $MSE_{WPerror}$ = Σ (y – person mean – cond mean + grand mean)^2) / df(48) = 7.00

Why error and person*age are the same thing given 2-occasion data



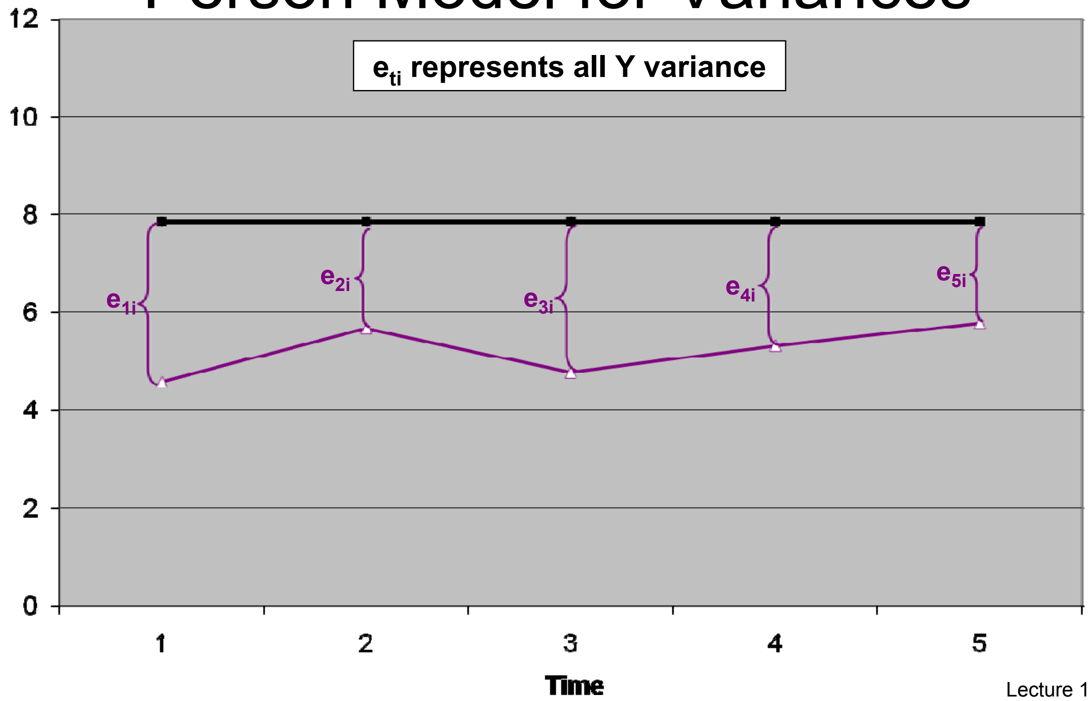
Lecture 1: Review of General Linear Models

- What is Multilevel Modeling?
- Between-Person vs. +Within-Person Models
- SPSS, SAS, and STATA GLM Examples
- **Summary of ANOVA Options for Longitudinal Data**

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



ANOVA for longitudinal data?

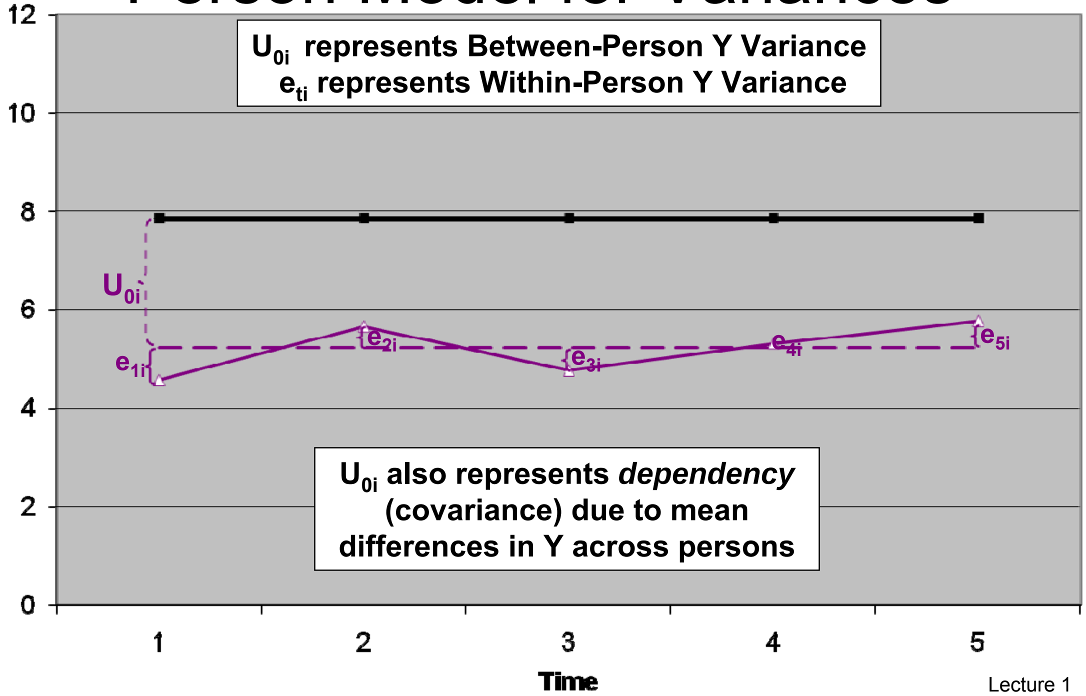
1. **Between-Groups Regression or ANOVA**
   – BP variance only (1 pile of e_{ti} only)

$$\begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & \sigma_e^2 \end{bmatrix}$$

   – 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



ANOVA for longitudinal data?

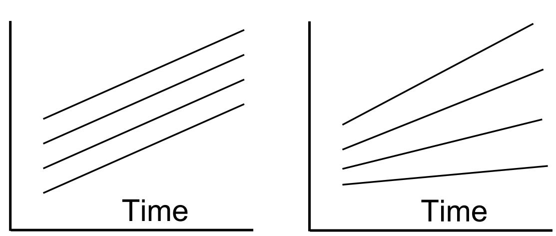
2. (a) Univariate Repeated Measures ANOVA: Var(e_{ti}) + Var(U_{0i})

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

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

$$\begin{bmatrix} \sigma_e^2+\tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \sigma_e^2+\tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \tau_{U_0}^2 & \sigma_e^2+\tau_{U_0}^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 & \sigma_e^2+\tau_{U_0}^2 \end{bmatrix}$$

– 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**



ANOVA for longitudinal data?

2. **(b) Univariate RM ANOVA with sphericity corrections**
   - Based on ε → how far off sphericity, (ranges 0-1, 1=spherical)
   - Applies an overall correction for model df based on estimated ε
   - Corrections for sphericity do not really solve the problem

3. **Multivariate Repeated Measures ANOVA**
   - Assumes nothing: all variances and covariances are estimated separately → "**Unstructured**"

It's not a model, it IS the data!

$$\begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22}^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33}^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44}^2 \end{bmatrix}$$

   - Because it can never it 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

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