# Introduction to Multilevel Models

- Topics:
  - **What is multilevel modeling?**
  - Concepts in longitudinal data
  - From between-person to within-person models
  - Kinds of ANOVAs for longitudinal data



# 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 Fluctuation 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 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 residuals are distributed and related across observations (persons, groups, time, etc.) → these relationships are called "dependency" and ***this is the primary way that multilevel models differ from general linear models*** *(e.g., regression)*



# Dimensions for Organizing Models

- Outcome type: General (normal) vs. General*ized* (not normal)
- Dimensions of sampling: One (so one variance term per outcome) vs. **Multiple** (so multiple variance terms per outcome) → **OUR WORLD**

- **General Linear Models:** conditionally normal outcome distribution, **fixed effects** (identity link; only one dimension of sampling)
- **General*ized* Linear Models:** **any conditional outcome distribution**, **fixed** effects through **link functions**, no random effects (one dimension)

Note: Least Squares is only for GLM

- **General Linear Mixed Models:** conditionally normal outcome distribution, **fixed and random effects** (identity link, but multiple sampling dimensions)
- **General*ized* Linear Mixed Models:** **any conditional outcome distribution**, **fixed and random effects** through **link functions** (multiple dimensions)

- "Linear" means the fixed effects predict the *link-transformed* DV in a linear combination of (effect*predictor) + (effect*predictor)…

# 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)
  - Persons (Level 1) within occasions (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 sources of correlation 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 soon)!**
   - Use SPSS Mixed, SAS Mixed, **Stata**, Mplus, R, HLM, MlwiN…
   - What's an intercept? What's a slope? What's a pile of variance?

# 1. Model Dependency

- <u>Sources of dependency depend on the sources of **variation** created by your sampling design:</u> residuals for outcomes from the same unit are likely to be related, which violates the GLM "independence" assumption

- **"Levels" for dependency** ="levels of random effects"
  - Sampling dimensions can be **nested**
    - e.g., time within person, person within group, school within country
  - If you can't figure out the direction of your nesting structure, odds are good you have a **crossed sampling design** instead
    - e.g., persons crossed with items, raters crossed with targets
  - To have a "level", there must be random outcome variation due to sampling that remains after including the model's fixed effects
    - e.g., treatment vs. control does not create another level of "group"



# Dependency comes from…

- Mean differences across sampling units (persons, groups)
  - Creates constant dependency over time (or persons)
  - Will be represented by a random intercept in our models

- Individual/group differences in effects of predictors
  - Longitudinal: individual differences in growth, stress reactivity
  - Clustered: group differences in slopes of person predictors
  - Creates non-constant dependency, the size of which depends on the value of the predictor at each occasion or for each person
  - Will be represented by random slopes in our models

- Longitudinal data: non-constant within-person correlation for unknown reasons (time-specific autocorrelation)
  - Can add other patterns of correlation as needed for this (AR, TOEP)

# Why care about dependency?

- In other words, what happens if we have the wrong model for the variances (assume independence instead)?
- **Validity of the tests of the predictors** depends on having the "most right" model for the variances
  - Estimates will usually be ok → come from model for the means
  - Standard errors (and thus *p*-values) can be inaccurate
- The sources of variation that exist in your outcome will dictate **what kinds of predictors** will be useful
  - Between-Person variation needs Between-Person predictors
  - Within-Person variation needs Within-Person predictors
  - Between-Group variation needs Between-Group predictors



# 2. Include categorical or continuous predictors at any level of analysis

- <u>ANOVA</u>: test differences among <u>discrete IV factor levels</u>
  - Between-Groups: Gender, Intervention Group, Age Groups
  - Within-Subjects (Repeated Measures): Condition, Time
  - Test main effects of continuous covariates (ANCOVA)
- <u>Regression</u>: test whether <u>slopes</u> relating predictors to outcomes are different from 0
  - Persons measured once, differ categorically or continuously on a set of time-invariant (person-level) covariates
- What if a predictor is assessed repeatedly (time-varying predictors) but can't be characterized by 'conditions'?
  - ANOVA or Regression won't work → need MLM

# 2. Include categorical or continuous predictors at any level of analysis

- Some things <u>don't change</u> over measurements…
  - Sex, Ethnicity
  
    → <u>Time-Invariant</u> Predictor = <u>Person</u> Level
- Some things <u>do change</u> over measurements…
  - Health Status, Stress Levels, Living Arrangements
  
    → <u>Time-Varying</u> Predictor = <u>Time</u> Level
- Some predictors might be measured at <u>higher levels</u>
  - Family SES, length of marriage, school size, country size
- Interactions between levels may be included, too
  - Does the effect of <u>health status</u> differ by <u>gender</u> and <u>SES</u>?

Level: Time Person Family



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

<u>RM ANOVA:</u> uses **multivariate** (wide) data structure:

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

<u>People</u> missing any data are excluded (data from ID 101 are not included at all)

<u>MLM:</u> uses **stacked** (long) data structure:

Only <u>rows</u> 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"…

# 4. You already know how!

- If you can do GLM, you can do MLM
  (*and if you can do generalized linear models, you can do generalized multilevel models, too*)

- How do you interpret an estimate for…
  - the intercept?
  - the effect of a continuous variable?
  - the effect of a categorical variable?
  - a variance component ("pile of variance")?



# Introduction to Multilevel Models

- Topics:
  - What is multilevel modeling?
  - **Concepts in longitudinal data**
  - From between-person to within-person models
  - Kinds of ANOVAs for longitudinal data

# Options for Longitudinal Models

- Although models and software are logically separate, longitudinal data can be analyzed via multiple analytic frameworks:
  - "Multilevel/Mixed Models"
    - Dependency over time, persons, groups, etc. is modeled via random effects (multivariate through "levels" using stacked/long data)
    - Builds on GLM, generalizes easier to additional levels of analysis
  - "Structural Equation Models"
    - Dependency over time *only* is modeled via latent variables (single-level analysis using multivariate/wide data)
    - Generalizes easier to broader analysis of latent constructs, mediation
  - Because random effects and latent variables are the same thing, many longitudinal models can be specified/estimated either way
    - And now "Multilevel Structural Equation Models" can do it all…



# Data Requirements for Our Models

- A useful outcome variable:
  - Has an interval scale*
    - A one-unit difference means the same thing across all scale points
    - In subscales, each contributing item has an equivalent scale
    - **Other kinds of outcomes can be analyzed using generalized multilevel models instead, but estimation is more challenging*
  - Has scores with the same meaning over observations
    - Includes meaning of construct
    - Includes how items relate to the scale
    - Implies measurement invariance
- FANCY MODELS CANNOT SAVE BADLY MEASURED VARIABLES OR CONFOUNDED RESEARCH DESIGNS.

# Requirements for Longitudinal Data

- Multiple OUTCOMES from the same sampling unit!
  - 2 is the minimum, but just 2 can lead to problems:
    - Only 1 kind of change is observable (1 difference)
    - Can't distinguish "real" individual differences in change from error
    - Repeated measures ANOVA is just fine for 2 observations
      - Necessary assumption of "sphericity" is satisfied with only 2 observations even if compound symmetry doesn't hold
  - 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 those individual differences
    - More items/stimuli →more power to show effects of differences between items/stimuli/conditions

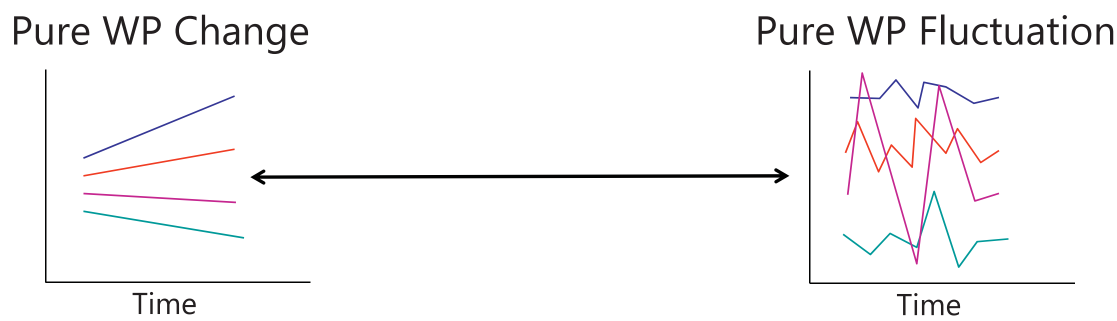


# Levels of Analysis in Longitudinal Data

- Between-Person (BP) Variation:
  - **Level-2** – "**INTER**-individual Differences" – Time-Invariant
  - All longitudinal studies begin as cross-sectional studies
- Within-Person (WP) Variation:
  - **Level-1** – "**INTRA**-individual Differences" – 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 BP and WP variation
  - BP = more/less than other people; WP = more/less than one's average
- I use "person" here, but level-2 can be anything that is measured repeatedly (like animals, schools, countries…)

# A Longitudinal Data Continuum

- **Within-Person Change**: Systematic change
  - Magnitude or direction of change can be different across individuals
  - "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 individual






# Introduction to Multilevel Models

- Topics:
  - What is multilevel modeling?
  - Concepts in longitudinal data
  - **From between-person to within-person models**
  - Kinds of ANOVAs for longitudinal data

# The Two Sides of a (BP) Model

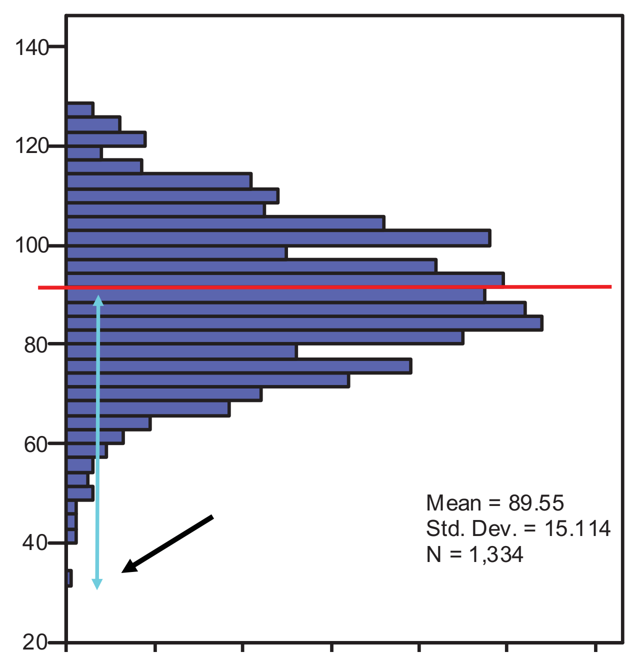
$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + e_i$$

Our focus today

- **Model for the Means (Predicted Values):**
  - Each person's expected (predicted) outcome is a weighted linear function of his/her values on X and Z (and here, their interaction), each measured once per person (i.e., this is a between-person model)
  - Estimated parameters are called fixed effects (here, $\beta_0$, $\beta_1$, $\beta_2$, and $\beta_3$)
- **Model for the Variance ("Piles" of Variance):**
  - $e_i \sim N(0, \sigma_e^2) \rightarrow$ ONE residual (unexplained) deviation
  - $e_i$ has a mean of 0 with some estimated constant variance $\sigma_e^2$, is normally distributed, is unrelated to X and Z, and is unrelated across people (across all observations, just people here)
  - **Estimated parameter is residual variance only in above BP model**



# An Empty Between-Person Model (i.e., Single-Level)




$$y_i = \beta_0 + e_i$$

**Filling in values:**

**32 = 90 + -58**

Y pred

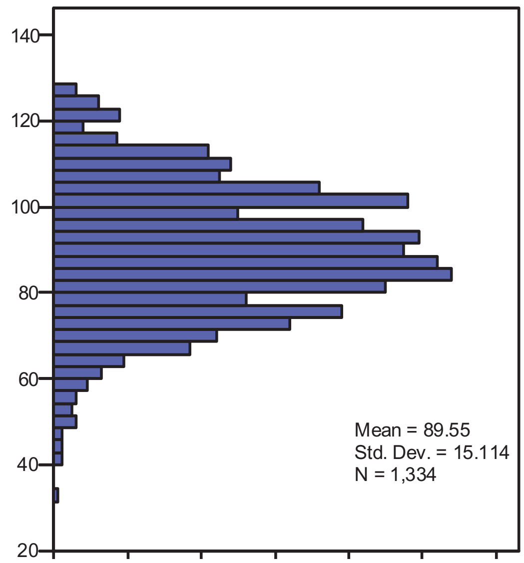
Model for the Means

Y Error Variance:

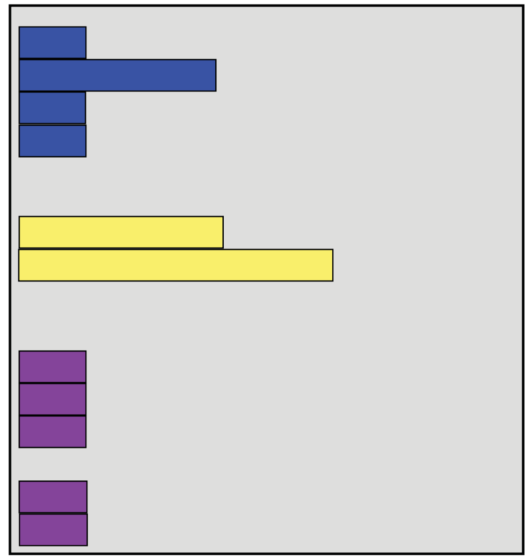
$$\frac{\Sigma (y - y_{pred})^2}{N - 1}$$

# Adding Within-Person Information… (i.e., to become a Multilevel Model)

Full Sample Distribution

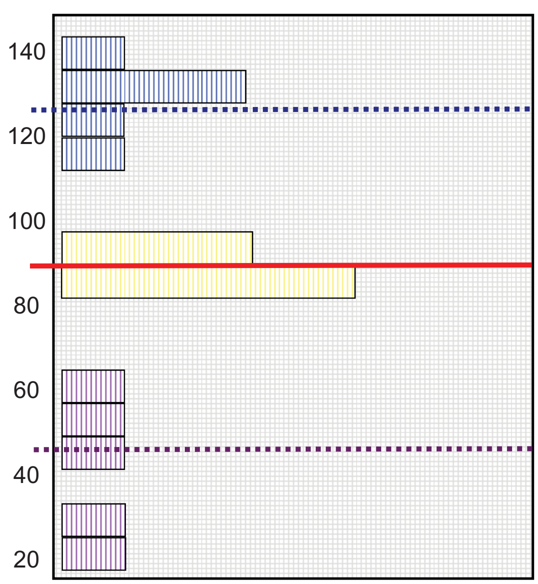



3 People, 5 Occasions each





# Empty +Within-Person Model




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

= Grand Mean

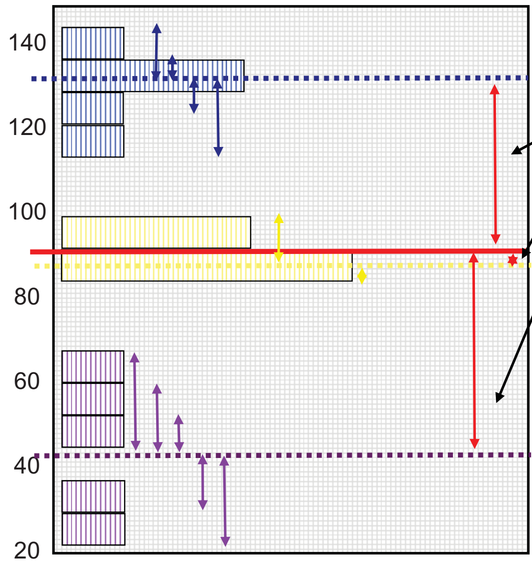
= Fixed Intercept

**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 (BP) Variance:**

- → Differences from **GRAND** mean
- → **INTER**-Individual Differences

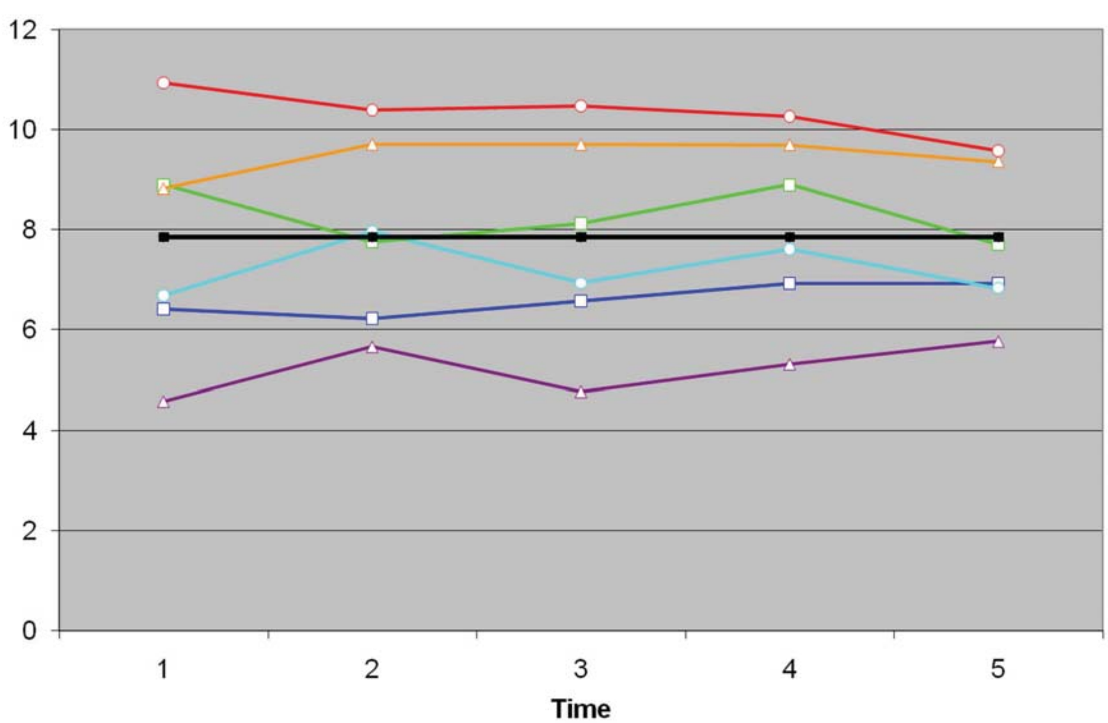
**Within-Person (WP) Variance:**

- → Differences from **OWN** mean
- → **INTRA**-Individual Differences
- → This part is only observable through longitudinal data.

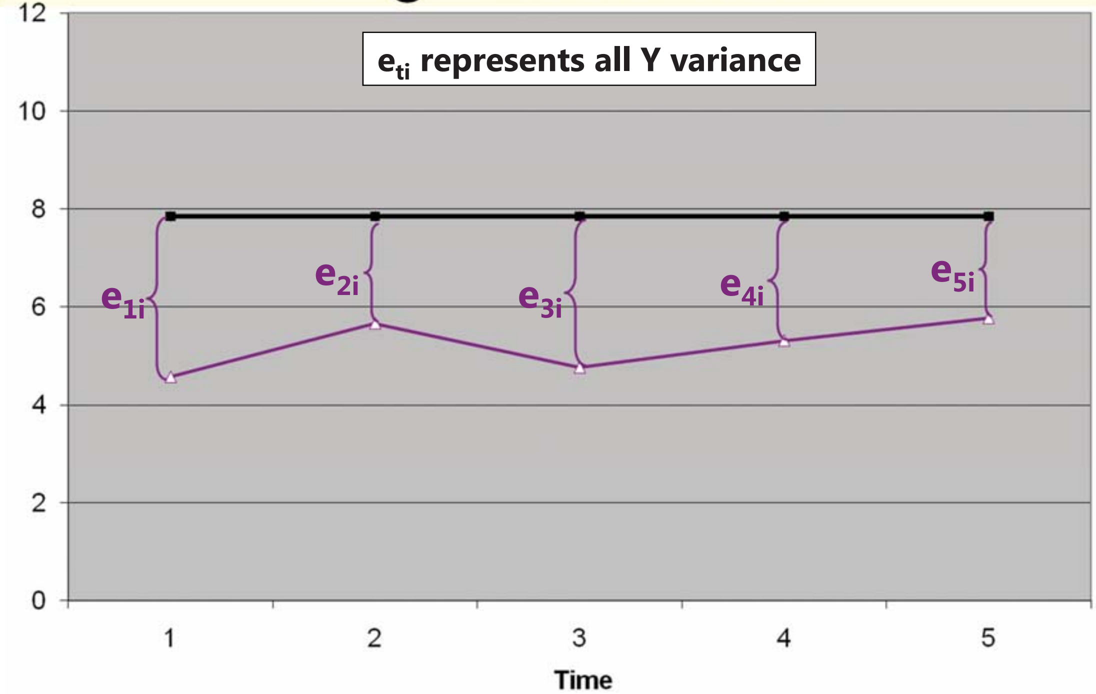
**Now we have 2 piles of variance in Y to predict.**



# Hypothetical Longitudinal Data

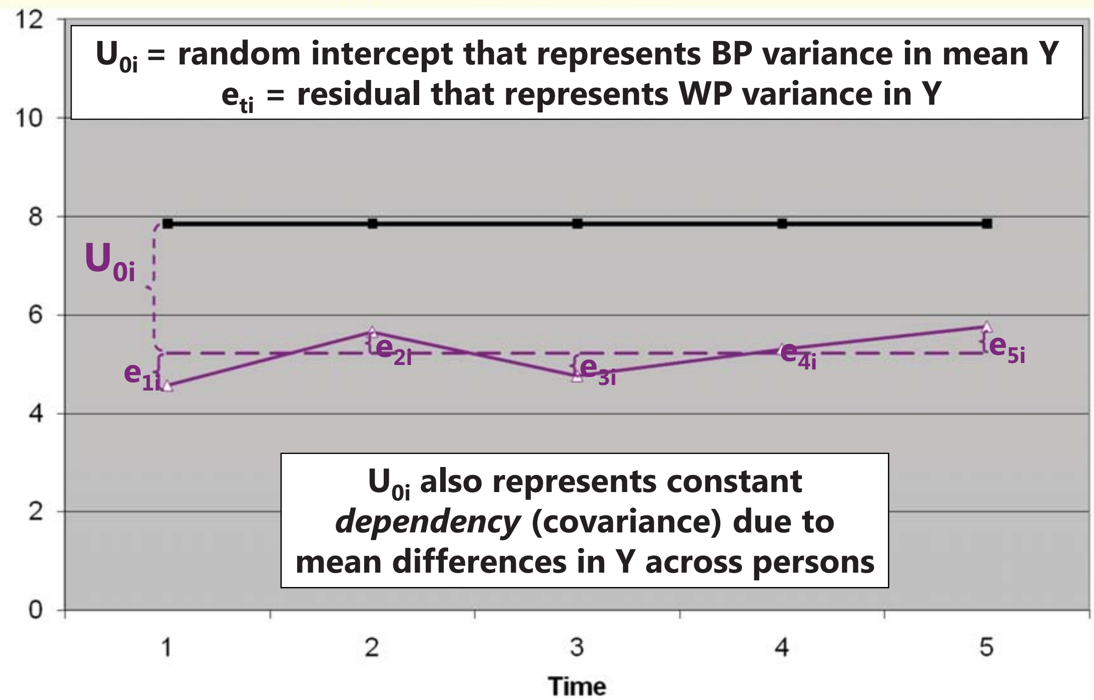


# "Error" in a BP Model for the Variance: Single-Level Model

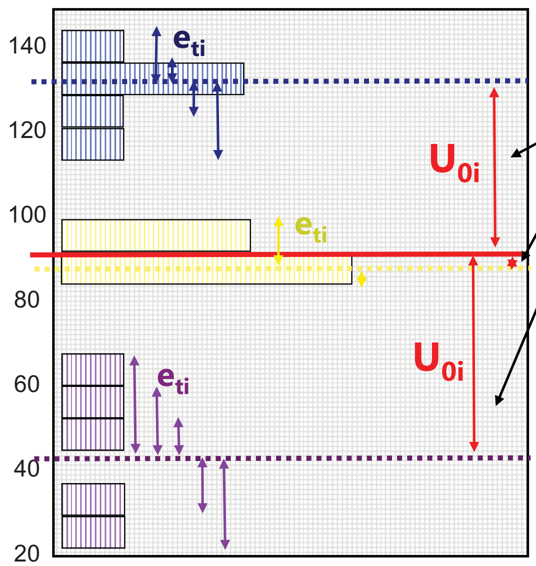





# "Error" in a +WP Model for the Variance: Multilevel Model



# Empty +Within-Person Model




Variance of Y → 2 sources:

**Level 2 Random Intercept Variance (of $U_{0i}$, as $\tau^2_{U_0}$):**

→ **Between**-Person Variance

→ Differences from **GRAND** mean

→ **INTER**-Individual Differences

**Level 1 Residual Variance (of $e_{ti}$, as $\sigma^2_e$):**

→ **Within**-Person Variance

→ Differences from **OWN** mean

→ **INTRA**-Individual Differences



# BP vs. +WP Empty Models

- Empty **Between-Person** Model (used for 1 occasion):

$$y_i = \beta_0 + e_i$$

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

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

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

  - $\beta_0$ = fixed intercept = grand mean
  - $U_{0i}$ = random intercept = individual deviation from GRAND mean
  - $e_{ti}$ = time-specific residual deviation from OWN mean

# Intraclass Correlation (ICC)

**Intraclass Correlation (ICC):**

$$\text{ICC} = \frac{\text{BP}}{\text{BP} + \text{WP}} = \frac{\text{Intercept Variance}}{\text{Intercept Variance} + \text{Residual Variance}}$$

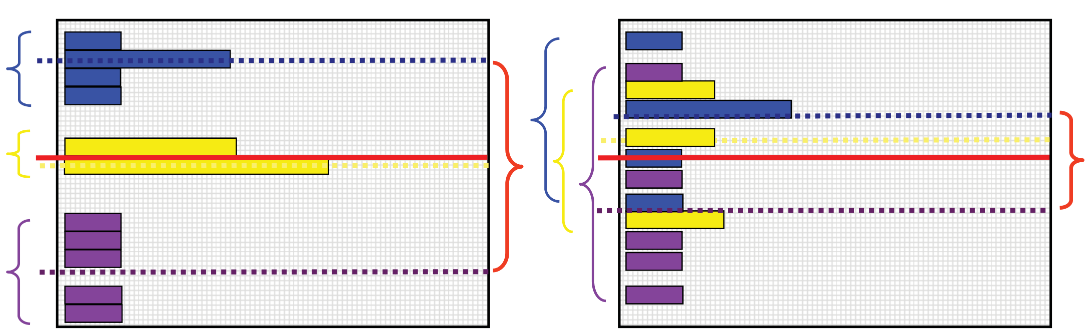
$$= \frac{\text{BP Variance in Mean Outcome}}{\text{Total Outcome Variance}} = \frac{\boldsymbol{\tau}^2_{U_0}}{\boldsymbol{\tau}^2_{U_0} + \boldsymbol{\sigma}^2_e}$$

- 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)**



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

Counter-Intuitive: Between-Person Variance is in the numerator, but the ICC is the correlation over time!



**ICC = BTW / BTW + within**

→ Large ICC

→ Large correlation over time

**ICC = btw / btw + WITHIN**

→ Small ICC

→ Small correlation over time

# BP and +WP Conditional Models

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

- Repeated Measures, **Within-Person** ANOVA: **2 PILES**
  - $\mathbf{y_{ti} = (\beta_0 + \beta_1 X_i + \beta_2 Z_i \ldots) + U_{0i} + e_{ti}}$
  - $\mathbf{U_{0i}}$ → A random intercept for differences in person means, assumed uncorrelated with equal variance across persons → "**BP (mean) variation**"= $\tau^2_{U_0}$ is now "leftover" after predictors
  - $\mathbf{e_{ti}}$ → A residual that represents remaining time-to-time variation, usually assumed uncorrelated with equal variance across observations (now, persons and time) → "**WP variation**" = $\sigma^2_e$ is also now "leftover" after predictors



# Introduction to Multilevel Models

- Topics:
  - What is multilevel modeling?
  - Concepts in longitudinal data
  - From between-person to within-person models
  - **Kinds of ANOVAs 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 due to least squares)
  - **Time-varying predictors** (covariates are BP predictors only)
- Each includes the same model for the means for 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!**
- These ANOVAs differ by what they predict for the correlation across outcomes from the same person in the model for the variances…
  - i.e., **how they "handle dependency"** due to persons, or what they says the variance and covariance of the $y_{ti}$ residuals should look like…



# 1. Between-Groups ANOVA

- **Uses $e_{ti}$ only** (total variance = a single variance term of $\sigma_e^2$)
- **Assumes no covariance** at all among observations from the same person: *Dependency? What dependency?*
- Will usually be **very, very wrong** for longitudinal data
  - WP effects tested against wrong residual variance (significance tests will often be way too conservative)
  - Will also tend to be wrong for clustered data, but less so *(because the correlation among persons from the same group is not as strong as the correlation among occasions from the same person)*
- Predicts a variance-covariance matrix over time (here, 4 occasions) like this, called "**Variance Components**":

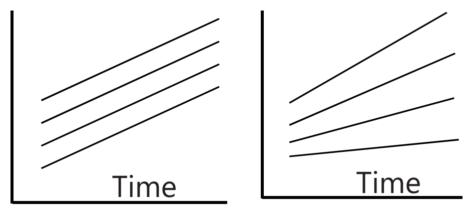
$$\begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & \sigma_e^2 \end{bmatrix}$$

# 2a. Univariate Repeated Measures

- Separates total variance into two sources:
  - **Between-Person** (mean differences due to $U_{0i}$, or $\boldsymbol{\tau^2_{U_0}}$)
  - **Within-Person** (remaining variance due to $e_{ti}$, or $\boldsymbol{\sigma^2_e}$)
- Predicts a variance-covariance matrix over time (here, 4 occasions) like this, called "**Compound Symmetry**":
  - **Mean differences from $U_{0i}$ are the only reason why occasions are correlated**

$$\begin{bmatrix} \sigma^2_e+\tau^2_{u_0} & \tau^2_{u_0} & \tau^2_{u_0} & \tau^2_{u_0} \\ \tau^2_{u_0} & \sigma^2_e+\tau^2_{u_0} & \tau^2_{u_0} & \tau^2_{u_0} \\ \tau^2_{u_0} & \tau^2_{u_0} & \sigma^2_e+\tau^2_{u_0} & \tau^2_{u_0} \\ \tau^2_{u_0} & \tau^2_{u_0} & \tau^2_{u_0} & \sigma^2_e+\tau^2_{u_0} \end{bmatrix}$$

- Will usually be at least somewhat wrong for longitudinal data
  - If people change at different rates, the variances and covariances over time have to change, too






# The Problem with Univariate RM ANOVA

- Univ. RM ANOVA $\left(\tau^2_{U_0}+\sigma^2_e\right)$ predicts **compound symmetry:**
  - All variances and all covariances are equal across occasions
  - In other words, the amount of error observed should be the same at any occasion, so a single, pooled error variance term makes sense
  - If not, tests of fixed effects may be biased (i.e., sometimes tested against too much or too little error, if error is not really constant over time)
  - **COMPOUND SYMMETRY RARELY FITS FOR LONGITUDINAL DATA**
- But to get the correct tests of the fixed effects, the data must only meet a less restrictive assumption of **sphericity**:
  - In English → **pairwise differences** between adjacent occasions have equal variance and covariance (satisfied by default with only 2 occasions)
  - If compound symmetry is satisfied, so is sphericity (but see above)
  - Significance test provided in ANOVA for where data meet sphericity assumption
  - **Other RM ANOVA approaches are used when sphericity fails…**

# The Other Repeated Measures ANOVAs…

- 2b. **Univariate RM ANOVA with sphericity corrections**
  - Based on ε → how far off sphericity (from 0-1, 1=spherical)
  - Applies an overall correction for model df based on estimated ε, but it doesn't really address the problem that data ≠ model

- 3. **Multivariate Repeated Measures ANOVA**
  - All variances and covariances are estimated separately over time (here, 4 occasions), called "**Unstructured**"—it's not a model, it IS the data reproduced directly:

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

  - Because it can never be wrong, UN can be useful for **complete and balanced longitudinal data** with few occasions (e.g., 2-4)
  - Parameters = $\frac{\#occasions * (\#occasions + 1)}{2}$ so can be hard to estimate
  - Unstructured can also be specified to include random intercept variance $\tau_{U_0}^2$
  - Every other model for the variances is nested within Unstructured (we can do model comparisons to see if all other models are NOT WORSE)



# 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)
  - BP ANOVA ($\sigma_e^2$ only) → assumes independence and constant variance over time
  - Univ. RM ANOVA $\left(\tau_{U_0}^2 + \sigma_e^2\right)$ → assumes constant variance and covariance
  - Multiv. RM ANOVA (whatever) → no assumptions; is a description, not a model

    there is no structure that shows up in a scalar equation (i.e., the way $U_{0i}$ + $e_{ti}$ does)

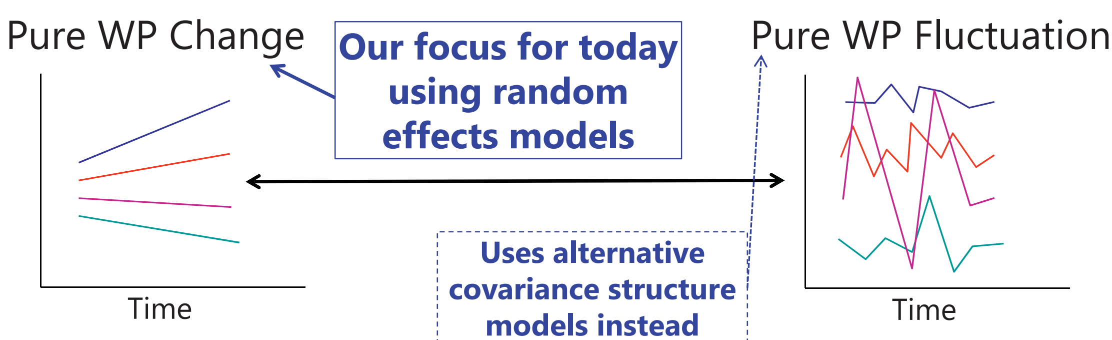
- **MLM will give us more flexibility in both parts of the model:**
  - Fixed effects that *predict* the pattern of means (polynomials, pieces)
  - Random intercepts and slopes and/or alternative covariance structures that *predict* intermediate patterns of variance and covariance over time

# Describing Within-Person Change in Longitudinal Data

- Topics:
  - **Multilevel modeling notation and terminology**
  - Fixed and random effects of linear time
  - Predicted variances and covariances from random slopes
  - Dependency and effect size in random effects models
  - Describing nonlinear change: polynomial, piecewise models, and truly nonlinear models
  - Fun with likelihood estimation and model comparisons



# Modeling Change vs. Fluctuation




**Model for the Means:**

- **WP Change** → describe pattern of *average* change (over "time")
- WP Fluctuation → *may* not need anything (if no systematic change)

**Model for the Variances:**

- **WP Change** → describe *individual differences* in change (random effects)
  → this allows variances and covariances to differ over time
- WP Fluctuation → describe pattern of variances and covariances over time

# The Big Picture of Longitudinal Data: Models for the Means

- What kind of change occurs on average over "time"? There are two baseline models to consider:
  - "**Empty**" → only a fixed intercept (predicts no change)
  - "**Saturated**" → all occasion mean differences from time 0
    (ANOVA model that uses # fixed effects= *n*)
    *** *may not be possible in unbalanced data*

Parsimony ← → Good fit

**Empty Model:**
Predicts NO change over time
**1 Fixed Effect**

In-between options:
polynomial slopes,
piecewise slopes,
nonlinear models…

**Saturated Means:**
Reproduces mean at each occasion
**# Fixed Effects = # Occasions**

***Name… that… Trajectory!***



# The Big Picture of Longitudinal Data: Models for the Variance

Parsimony ← → Good fit

Compound Symmetry (CS)

$$\begin{bmatrix} \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e \end{bmatrix}$$

Univariate RM ANOVA

Most useful model: likely somewhere in between!

**NAME …THAT … STRUCTURE!**

Unstructured (UN)

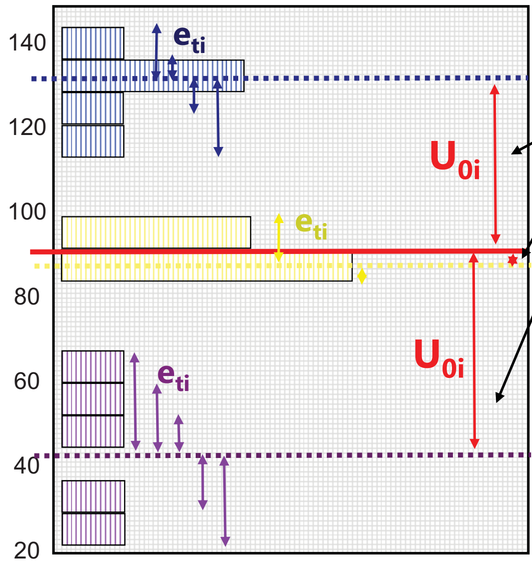
$$\begin{bmatrix} \sigma^2_{T1} & \sigma_{T12} & \sigma_{T13} & \sigma_{T14} \\ \sigma_{T21} & \sigma^2_{T2} & \sigma_{T23} & \sigma_{T24} \\ \sigma_{T31} & \sigma_{T32} & \sigma^2_{T3} & \sigma_{T43} \\ \sigma_{T41} & \sigma_{T42} & \sigma_{T43} & \sigma^2_{T4} \end{bmatrix}$$

Multivariate RM ANOVA

***What is the pattern of variance and covariance over time?***

CS and UN are just two of the many, many options available within MLM, including ***random effects models*** (for change) and ***alternative covariance structure models*** (for fluctuation).

# Empty +Within-Person Model




Variance of Y → 2 sources:

**Level 2 Random Intercept Variance (of $U_{0i}$, as $\tau^2_{U_0}$):**

→ **Between**-Person Variance

→ Differences from **GRAND** mean

→ **INTER**-Individual Differences

**Level 1 Residual Variance (of $e_{ti}$, as $\sigma^2_e$):**

→ **Within**-Person Variance

→ Differences from **OWN** mean

→ **INTRA**-Individual Differences



# Empty Means, Random Intercept Model

GLM Empty Model:

- $y_i = \beta_0 + e_i$

MLM Empty Model:

- Level 1:

  $y_{ti} = \beta_{0i} + e_{ti}$

- Level 2:

  $\beta_{0i} = \gamma_{00} + U_{0i}$

**Residual = time-specific deviation from individual's predicted outcome**

**Fixed Intercept =grand mean (because no predictors yet)**

**Random Intercept = individual-specific deviation from predicted intercept**

**3** Total Parameters:

**Model for the Means (1):**

- Fixed Intercept $\gamma_{00}$

**Model for the Variance (2):**

- Level-1 Variance of $e_{ti}$ → $\sigma^2_e$
- Level-2 Variance of $U_{0i}$ → $\tau^2_{U_0}$

**Composite equation:**

$y_{ti} = (\gamma_{00} + U_{0i}) + e_{ti}$

## Saturated Means, Random Intercept Model

- Although rarely shown this way, a saturated means, random intercept model would be represented as a multilevel model like this (for *n* = 4 here, in which the time predictors are dummy codes to distinguish each occasion from time 0):
- Level 1:
  $$y_{ti} = \beta_{0i} + \beta_{1i}(Time1_{ti}) + \beta_{2i}(Time2_{ti}) + \beta_{3i}(Time3_{ti}) + e_{ti}$$
- Level 2:
  $$\beta_{0i} = \gamma_{00} + U_{0i}$$
  $$\beta_{1i} = \gamma_{10}$$
  $$\beta_{2i} = \gamma_{20}$$
  $$\beta_{3i} = \gamma_{30}$$

Composite equation (6 parameters):
$$y_{ti} = \gamma_{00} + \gamma_{10}(Time1_{ti}) + \gamma_{20}(Time2_{ti}) + \gamma_{30}(Time3_{ti}) + U_{0i} + e_{ti}$$

This model is also known as **univariate repeated measures ANOVA**. Although the means are perfectly predicted, the random intercept assumes parallel growth (and equal variance/covariance over time).



# Describing Within-Person Change in Longitudinal Data

- Topics:
  - Multilevel modeling notation and terminology
  - **Fixed and random effects of linear time**
  - Predicted variances and covariances from random slopes
  - Dependency and effect size in random effects models
  - Describing nonlinear change: polynomial, piecewise models, and truly nonlinear models
  - Fun with likelihood estimation and model comparisons

# Augmenting the empty means, random intercept model with *time*

- 2 questions about the possible effects of *time*:

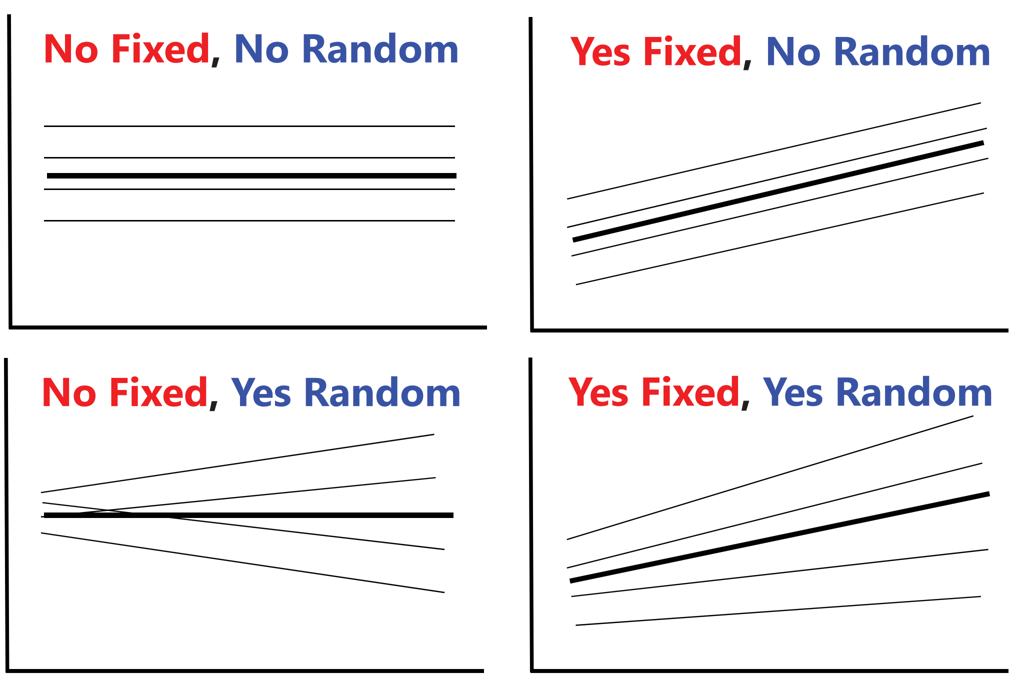
1. **Is there an effect of time on average?**
   - If the line describing the sample means not flat?
   - Significant **FIXED** effect of time

2. **Does the average effect of time vary across individuals?**
   - Does each individual need his or her own line?
   - Significant **RANDOM** effect of time



# Fixed and Random Effects of Time

(Note: The intercept is random in every figure)



# Fixed Linear Time, Random Intercept Model

(4 total parameters: effect of time is FIXED only)

Multilevel Model

**Residual** = time-specific deviation from individual's predicted outcome → estimated variance of $\sigma_e^2$

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}$

**Fixed Intercept** = predicted mean outcome at time 0

**Fixed Linear Time Slope** = predicted mean rate of change per unit time

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

**Random Intercept** = individual-specific deviation from fixed intercept → estimated variance of $\tau_{U0}^2$

Composite Model

$$y_{ti} = \underbrace{(\gamma_{00} + U_{0i})}_{\beta_{0i}} + \underbrace{(\gamma_{10})}_{\beta_{1i}}(Time_{ti}) + e_{ti}$$

Because the effect of time is **fixed**, everyone is predicted to change at exactly the same rate.



# Random Intercept Models Imply…

- **People differ from each other systematically in only ONE way**—in intercept ($U_{0i}$), which implies **ONE kind of BP variance**, which translates to **ONE source of person dependency** (covariance or correlation in the outcomes from the same person)
- If so, after controlling for BP intercept differences (by estimating the variance of $U_{0i}$ as $\tau_{U0}^2$ in the **G** matrix), the **$e_{ti}$ residuals** (whose variance and covariance are estimated in the **R** matrix) should be **uncorrelated with homogeneous variance across time**, as shown:

Level-2 **G** matrix: RANDOM TYPE=UN

$$\begin{bmatrix} \tau_{U_0}^2 \end{bmatrix}$$

Level-1 **R** matrix: REPEATED TYPE=VC

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

**G** and **R** matrices combine to create a total **V** matrix with CS pattern

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

# Matrices in a Random Intercept Model

**Total predicted data matrix is called V matrix, created from the G [TYPE=UN] and R [TYPE=VC] matrices as follows:**

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}\tau^2_{U_0}\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}\sigma^2_e & 0 & 0 & 0\\ 0 & \sigma^2_e & 0 & 0\\ 0 & 0 & \sigma^2_e & 0\\ 0 & 0 & 0 & \sigma^2_e\end{bmatrix} = \begin{bmatrix}\tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e\end{bmatrix}$$

**VCORR then provides the intraclass correlation, calculated as:**
**ICC = $\tau^2_{U_0} / (\tau^2_{U_0} + \sigma^2_e)$**

$$\begin{bmatrix}1 & ICC & ICC & ICC\\ ICC & 1 & ICC & ICC\\ ICC & ICC & 1 & ICC\\ ICC & ICC & ICC & 1\end{bmatrix}$$

assumes a constant correlation over time

**For any random effects model:**

**G matrix** = BP variances/covariances

**R matrix** = WP variances/covariances

**Z matrix** = values of predictors with random effects (just intercept here), which can vary per person

**V matrix** = Total variance/covariance



# Summary so far…

- Regardless of what kind of model for the means you have…
  - Empty means = 1 fixed intercept that predicts no change
  - Saturated means = 1 fixed intercept + *n*−1 fixed effects for mean differences that perfectly predict the means over time
    - Is a description, not a model, and may not be possible with unbalanced time
  - Fixed linear time = 1 fixed intercept, 1 fixed linear time slope that predicts linear average change across time
    - Is a model that works with balanced or unbalanced time
    - May cause an increase in the random intercept variance by explaining residual variance
- A random intercept model…
  - Predicts constant total variance and covariance over time in **V** using **G**
    - Should be possible in balanced or unbalanced data
  - Still has residual variance (always there via default **R** matrix TYPE=VC)
- Now we'll see what happens when adding other kinds of random effects, such as a random linear effect of time…

# Random Linear Time Model (6 total parameters)

## Multilevel Model

**Residual** = time-specific deviation from individual's predicted outcome → estimated variance of $\sigma_e^2$

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}$

**Fixed Intercept** = predicted mean outcome at time 0

**Fixed Linear Time Slope** = predicted mean rate of change per unit time

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

**Random Intercept** = individual-specific deviation from fixed intercept at time 0 → estimated variance of $\tau_{U_0}^2$

**Random Linear Time Slope** = individual-specific deviation from fixed linear time slope → estimated variance of $\tau_{U_1}^2$

Also has an estimated **covariance** of random intercepts and slopes of $\tau_{U_{01}}$

## Composite Model

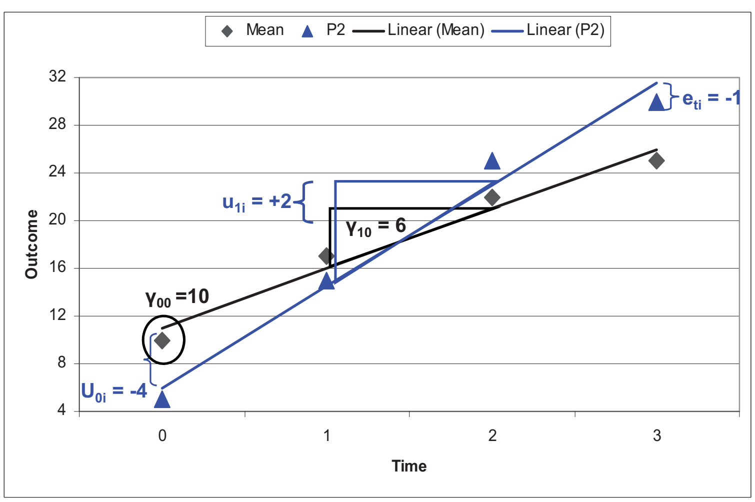
$$y_{ti} = \underbrace{(\gamma_{00} + U_{0i})}_{\beta_{0i}} + \underbrace{(\gamma_{10} + U_{1i})}_{\beta_{1i}}(Time_{ti}) + e_{ti}$$



# Random Linear Time Model

$$y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{1i})(Time_{ti}) + e_{ti}$$

- $\gamma_{00}$: Fixed Intercept
- $U_{0i}$: Random Intercept Deviation
- $\gamma_{10}$: Fixed Slope
- $U_{1i}$: Random Slope Deviation
- $e_{ti}$: error for person i at time t



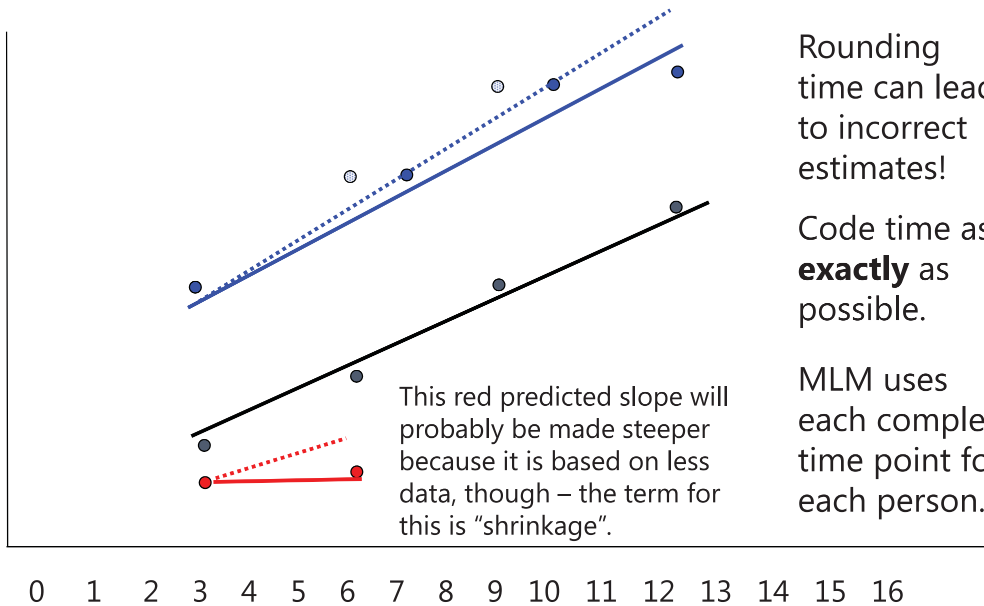

**6 Parameters:**

2 Fixed Effects:
$\gamma_{00}$ Intercept, $\gamma_{10}$ Slope

2 Random Effects Variances:
$U_{0i}$ Intercept Variance $= \tau_{U_0}^2$
$U_{1i}$ Slope Variance $= \tau_{U_1}^2$
Int-Slope Covariance $= \tau_{U_{01}}$

1 $e_{ti}$ Residual Variance $= \sigma_e^2$

# Unbalanced Time → Different time occasions across persons? No problem!




Rounding time can lead to incorrect estimates!

Code time as **exactly** as possible.

MLM uses each complete time point for each person.



# Summary: Sequential Models for Effects of Time

Level 1: $\mathbf{y_{ti}} = \boldsymbol{\beta_{0i}} + e_{ti}$

Level 2: $\boldsymbol{\beta_{0i}} = \gamma_{00} + U_{0i}$

Composite: $\mathbf{y_{ti}} = \gamma_{00} + U_{0i} + e_{ti}$

Empty Means, Random Intercept Model: 3 parms = $\gamma_{00}, \sigma^2_e, \tau^2_{U_0}$

---

Level 1: $\mathbf{y_{ti}} = \boldsymbol{\beta_{0i}} + \boldsymbol{\beta_{1i}}(\mathbf{Time_{ti}}) + e_{ti}$

Level 2: $\boldsymbol{\beta_{0i}} = \gamma_{00} + U_{0i}$

$\boldsymbol{\beta_{1i}} = \gamma_{10}$

Composite: $\mathbf{y_{ti}} = (\gamma_{00} + U_{0i}) + \gamma_{10}(\mathbf{Time_{ti}}) + e_{ti}$

Fixed Linear Time, Random Intercept Model: 4 parms = $\gamma_{00}, \gamma_{10}, \sigma^2_e, \tau^2_{U_0}$

---

Level 1: $\mathbf{y_{ti}} = \boldsymbol{\beta_{0i}} + \boldsymbol{\beta_{1i}}(\mathbf{Time_{ti}}) + e_{ti}$

Level 2: $\boldsymbol{\beta_{0i}} = \gamma_{00} + U_{0i}$

$\boldsymbol{\beta_{1i}} = \gamma_{10} + U_{1i}$

Composite: $\mathbf{y_{ti}} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{0i})(\mathbf{Time_{ti}}) + e_{ti}$

Random Linear Time Model: 6 parms = $\gamma_{00}, \gamma_{10}, \sigma^2_e, \tau^2_{U_0}, \tau^2_{U_1}, \tau_{U_{01}}$ (→ cov of $U_{0i}$ and $U_{1i}$)

# Describing Within-Person Change in Longitudinal Data

- Topics:
  - Multilevel modeling notation and terminology
  - Fixed and random effects of linear time
  - **Predicted variances and covariances from random slopes**
  - Dependency and effect size in random effects models
  - Describing nonlinear change: polynomial, piecewise models, and truly nonlinear models
  - Fun with likelihood estimation and model comparisons



# Random Linear Time Models Imply:

- **People differ from each other systematically in TWO ways**—in intercept ($U_{0i}$) and slope ($U_{1i}$), which implies **TWO kinds of BP variance**, which translates to **TWO sources of person dependency** (covariance or correlation in the outcomes from the same person)
- If so, after controlling for both BP intercept and slope differences (by estimating the $\tau^2_{U_0}$ and $\tau^2_{U_1}$ variances in the **G** matrix), the **$e_{ti}$ residuals** (whose variance and covariance are estimated in the **R** matrix) should be **uncorrelated with homogeneous variance across time**, as shown:

Level-2 **G** matrix: RANDOM TYPE=UN

$$\begin{bmatrix} \tau^2_{U_0} & \tau_{U_{10}} \\ \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix}$$

Level-1 **R** matrix: REPEATED TYPE=VC

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

**G** and **R** combine to create a total **V** matrix whose per-person structure depends on the specific time occasions each person has (very flexible for unbalanced time)

# Random Linear Time Model
## (6 total parameters: effect of time is now RANDOM)

- How the model predicts each element of the **V** matrix:

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}$

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

Composite Model: $y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{0i})(Time_{ti}) + e_{ti}$

**Predicted *Time-Specific* Variance:**

$$\begin{aligned}
Var[y_{ti}] &= Var[(\gamma_{00}+U_{0i})+(\gamma_{10}+U_{1i})(Time_i)+e_{ti}] \\
&= Var[(U_{0i})+(U_{1i}*Time_i)+e_{ti}] \\
&= \{Var(U_{0i})\}+\{Var(U_{1i}*Time_i)\}+\{2*Cov(U_{0i},U_{1i}*Time_i)\}+\{Var(e_{ti})\} \\
&= \{Var(U_{0i})\}+\{Time_i^2*Var(U_{1i})\}+\{2*Time_i*Cov(U_{0i},U_{1i})\}+\{Var(e_{ti})\} \\
&= \{\tau_{U_0}^2\}+\boxed{\{Time_i^2*\tau_{U_1}^2\}+\{2*Time_i*\tau_{U_{01}}\}}+\{\sigma_e^2\}
\end{aligned}$$



# Random Linear Time Model
## (6 total parameters: effect of time is now RANDOM)

- How the model predicts each element of the **V** matrix:

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}$

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

Composite Model: $y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{0i})(Time_{ti}) + e_{ti}$

**Predicted *Time-Specific* Covariances (Time A with Time B):**

$$\begin{aligned}
Cov[y_{Ai},y_{Bi}] &= Cov[\{(\gamma_{00}+U_{0i})+(\gamma_{10}+U_{1i})(A_i)+e_{Ai}\},\{(\gamma_{00}+U_{0i})+(\gamma_{10}+U_{1i})(B_i)+e_{Bi}\}] \\
&= Cov[\{U_{0i}+(U_{1i}A_i)\},\{U_{0i}+(U_{1i}B_i)\}] \\
&= Cov[U_{0i},U_{0i}]+Cov[U_{0i},U_{1i}B_i]+Cov[U_{0i},U_{1i}A_i]+Cov[U_{1i}A_i,U_{1i}B_i] \\
&= \{Var(U_{0i})\}+\{(A_i+B_i)*Cov(U_{0i},U_{1i})\}+\{(A_iB_i)Var(U_{1i})\} \\
&= \{\tau_{U_0}^2\}+\boxed{\{(A_i+B_i)\tau_{U_{01}}\}+\{(A_iB_i)\tau_{U_1}^2\}}
\end{aligned}$$

# Random Linear Time Model
## (6 total parameters: effect of time is now RANDOM)

- Scalar "mixed" model equation per person:

$$\mathbf{Y}_i = \mathbf{X}_i * \boldsymbol{\gamma} + \mathbf{Z}_i * \mathbf{U}_i + \mathbf{E}_i$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} U_{0i} \\ U_{1i} \end{bmatrix} + \begin{bmatrix} e_{0i} \\ e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} \gamma_{00}+\gamma_{10}(0) \\ \gamma_{00}+\gamma_{10}(1) \\ \gamma_{00}+\gamma_{10}(2) \\ \gamma_{00}+\gamma_{10}(3) \end{bmatrix} + \begin{bmatrix} U_{0i}+U_{1i}(0) \\ U_{0i}+U_{1i}(1) \\ U_{0i}+U_{1i}(2) \\ U_{0i}+U_{1i}(3) \end{bmatrix} + \begin{bmatrix} e_{0i} \\ e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} \gamma_{00}+\gamma_{10}(0)+U_{0i}+U_{1i}(0)+e_{0i} \\ \gamma_{00}+\gamma_{10}(1)+U_{0i}+U_{1i}(1)+e_{1i} \\ \gamma_{00}+\gamma_{10}(2)+U_{0i}+U_{1i}(2)+e_{2i} \\ \gamma_{00}+\gamma_{10}(3)+U_{0i}+U_{1i}(3)+e_{3i} \end{bmatrix}$$

$\mathbf{X}_i$ = *n* x *k* values of **predictors with fixed effects**, so can differ per person (*k* = 2: intercept, linear time)

$\boldsymbol{\gamma}$ = *k* x 1 estimated **fixed effects**, so will be the same for all persons ($\gamma_{00}$ = intercept, $\gamma_{10}$ = linear time)

$\mathbf{Z}_i$ = *n* x *u* values of **predictors with random effects**, so can differ per person (*u* = 2: intercept, linear time)

$\mathbf{U}_i$ = *u* x 2 estimated individual **random effects**, so can differ per person

$\mathbf{E}_i$ = *n* x *n* time-specific residuals, so can differ per person



# Random Linear Time Model
## (6 total parameters: effect of time is now RANDOM)

- Predicted total variances and covariances per person:

$$\mathbf{V}_i = \mathbf{Z}_i * \mathbf{G}_i * \mathbf{Z}_i^T + \mathbf{R}_i$$

$$\mathbf{V}_i = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} \\ \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

$\mathbf{V}_i$ matrix: Variance$[y_{time}]$

$$= \tau^2_{U_0} + \left[(\text{time})^2 \tau^2_{U_1}\right] + \left[2(\text{time})\tau_{U_{01}}\right] + \sigma^2_e$$

$\mathbf{V}_i$ matrix = complicated ☺

$\mathbf{V}_i$ matrix: Covariance$[y_A, y_B]$

$$= \tau^2_{U_0} + \left[(A+B)\tau_{U_{01}}\right] + \left[(AB)\tau^2_{U_1}\right]$$

$\mathbf{Z}_i$ = *n* x *u* values of **predictors with random effects**, so can differ per person (*u* = 2: int., time slope)

$\mathbf{Z}_i^T$ = *u* x *n* values of predictors with random effects (just $\mathbf{Z}_i$ transposed)

$\mathbf{G}_i$ = *u* x *u* estimated **random effects variances and covariances**, so will be the same for all persons ($\tau^2_{U0}$ = int. var., $\tau^2_{U1}$ = slope var.)

$\mathbf{R}_i$ = *n* x *n* **time-specific residual variances and covariances**, so will be same for all persons (here, just diagonal $\sigma^2_e$)

# Building **V** across persons: Random Linear Time Model

- **V** for two persons with **unbalanced time** observations:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R}$$

$$\mathbf{V} = \begin{bmatrix} 1 & 0.0 & 0 & 0 \\ 1 & 1.0 & 0 & 0 \\ 1 & 2.0 & 0 & 0 \\ 1 & 3.0 & 0 & 0 \\ 0 & 0 & 1 & 0.2 \\ 0 & 0 & 1 & 1.4 \\ 0 & 0 & 1 & 2.3 \\ 0 & 0 & 1 & 3.5 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} & 0 & 0 \\ \tau_{U_{01}} & \tau^2_{U_1} & 0 & 0 \\ 0 & 0 & \tau^2_{U_0} & \tau_{U_{01}} \\ 0 & 0 & \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0.0 & 1.0 & 2.0 & 3.0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0.2 & 1.4 & 2.3 & 3.5 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

- The giant combined **V** matrix across persons is how the multilevel or mixed model is actually estimated
- Known as "**block diagonal**" structure → predictions are given for each person, but 0's are given for the elements that describe relationships between persons (because persons are supposed to be independent here!)



# Building **V** across persons: Random Linear Time Model

- **V** for two persons also with **different *n*** per person:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R}$$

$$\mathbf{V} = \begin{bmatrix} 1 & 0.0 & 0 & 0 \\ 1 & 1.0 & 0 & 0 \\ 1 & 2.0 & 0 & 0 \\ 1 & 3.0 & 0 & 0 \\ 0 & 0 & 1 & 0.2 \\ 0 & 0 & 1 & 1.4 \\ 0 & 0 & 1 & 3.5 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} & 0 & 0 \\ \tau_{U_{01}} & \tau^2_{U_1} & 0 & 0 \\ 0 & 0 & \tau^2_{U_0} & \tau_{U_{01}} \\ 0 & 0 & \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0.0 & 1.0 & 2.0 & 3.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0.2 & 1.4 & 3.5 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

- The "block diagonal" does not need to be the same size or contain the same time observations per person…
- **R** matrix can also include non-0 covariance or differential residual variance across time (as in ACS models), although the models based on the idea of a "lag" won't work for unbalanced or unequal-interval time

# G, R, and V: The Take-Home Point

- 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 via **Z** 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
    - Can allow variance and covariance due to other predictors, too

Compound Symmetry (CS)

$$\begin{bmatrix} \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e \end{bmatrix}$$

Random effects models use **G** and **R** to predict something in-between!

Unstructured (UN)

$$\begin{bmatrix} \sigma^2_{T1} & \sigma_{T12} & \sigma_{T13} & \sigma_{T14} \\ \sigma_{T21} & \sigma^2_{T2} & \sigma_{T23} & \sigma_{T24} \\ \sigma_{T31} & \sigma_{T32} & \sigma^2_{T3} & \sigma_{T43} \\ \sigma_{T41} & \sigma_{T42} & \sigma_{T43} & \sigma^2_{T4} \end{bmatrix}$$

**Parsimony** ← → **Good fit**

Univariate RM ANOVA — Multivariate RM ANOVA



# Describing Within-Person Change in Longitudinal Data

- Topics:
  - Multilevel modeling notation and terminology
  - Fixed and random effects of linear time
  - Predicted variances and covariances from random slopes
  - **Dependency and effect size in random effects models**
  - Describing nonlinear change: polynomial, piecewise models, and truly nonlinear models
  - Fun with likelihood estimation and model comparisons

# How MLM "Handles" Dependency

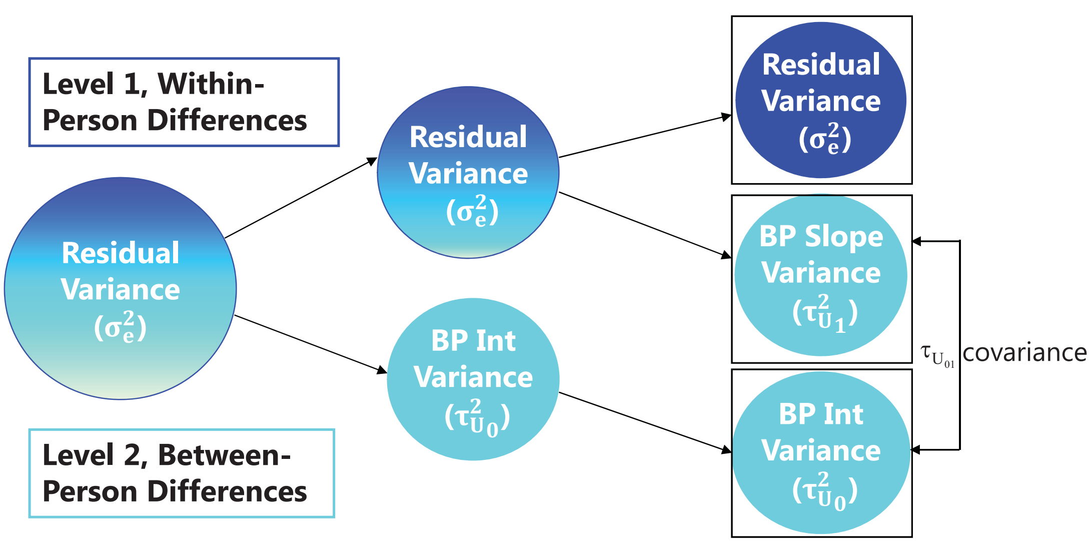
- Common description of the purpose of MLM is that it "addresses" or "handles" correlated (dependent) data…
- But where does this correlation come from?
  3 places (here, an example with health as an outcome):

1. *Mean differences across persons*
   - Some people are just healthier than others (at every time point)
   - This is what a random intercept is for
2. *Differences in effects of predictors across persons*
   - Does *time* (or *stress*) affect health more in some persons than others?
   - This is what random slopes are for
3. Non-constant within-person correlation for unknown reasons
   - Occasions closer together may just be more related
   - This is what ACS models are for



# MLM "Handles" Dependency

- Where does each kind of person dependency go? Into a new random effects variance component (or "pile" of variance):



# Piles of Variance

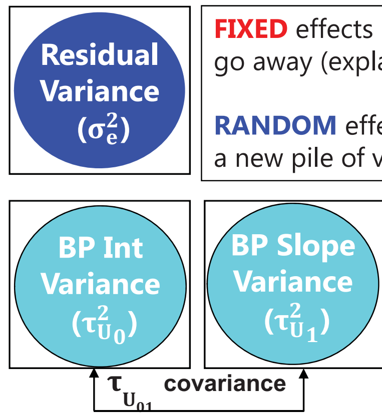
- By adding a random slope, we **carve up** our total variance into 3 piles:
  - BP (error) variance around intercept
  - BP (error) variance around slope
  - WP (error) residual variance

  } These 2 piles are 1 pile of "error variance" in Univ. RM ANOVA

- **But making piles does NOT make error variance go away…**

**Level 1 (one source of) <u>Within-Person Variation</u>: gets accounted for by time-level predictors**

**Level 2 (two sources of) <u>Between-Person Variation</u>: gets accounted for by person-level predictors**




**FIXED** effects make variance go away (explain variance).

**RANDOM** effects just make a new pile of variance.



# Fixed vs. Random Effects of Persons

- Person dependency: via **fixed effects in the model for the means** or via **random effects in the model for the variance**?
  - Individual intercept differences can be included as:
    - **N-1 person dummy code fixed main effects OR 1 random $U_{0i}$**
  - Individual time slope differences can be included as:
    - **N-1*time person dummy code interactions OR 1 random $U_{1i}$*$time_{ti}$**
  - Either approach would appropriately control for dependency (fixed effects are used in some programs that 'control' SEs for sampling)
- Two important advantages of **random effects**:
  - <u>Quantification</u>: Direct measure of how much of the outcome variance is due to person differences (in intercept or in effects of predictors)
  - <u>Prediction</u>: Person differences (main effects and effects of time) then become predictable quantities – this can't happen using fixed effects
  - **Summary: Random effects give you *predictable* control of dependency**

# Explained Variance from Fixed Linear Time

- Most common measure of effect size in MLM is Pseudo-$R^2$
  - Is supposed to be variance accounted for by predictors
  - Multiple piles of variance mean multiple possible values of pseudo $R^2$ (can be calculated per variance component or per model level)
  - A fixed linear effect of time will reduce level-1 residual variance $\sigma_e^2$ in **R**
  - By how much is the residual variance $\sigma_e^2$ reduced?

$$\text{Pseudo } R_e^2 = \frac{\text{residual variance}_{\text{fewer}} - \text{residual variance}_{\text{more}}}{\text{residual variance}_{\text{fewer}}}$$

  - If time varies between persons, then level-2 random intercept variance $\tau_{U_0}^2$ in **G** may also be reduced:

$$\text{Pseudo } R_{U0}^2 = \frac{\text{random intercept variance}_{\text{fewer}} - \text{random intercept variance}_{\text{more}}}{\text{random intercept variance}_{\text{fewer}}}$$

  - But you are likely to see a (net) INCREASE in $\tau_{U_0}^2$ instead.... Here's why:



# Increases in Random Intercept Variance

- Level-2 random intercept variance $\tau_{U_0}^2$ will often increase as a consequence of reducing level-1 residual variance $\sigma_e^2$
- Observed level-2 $\tau_{U_0}^2$ is NOT just between-person variance
  - Also has a small part of within-person variance (level-1 $\sigma_e^2$), or: **Observed $\tau_{U_0}^2$ = True $\tau_{U_0}^2$ + ($\sigma_e^2/n$)**
    - As *n* occasions increases, bias of level-1 $\sigma_e^2$ is minimized
  - Likelihood-based estimates of "true" $\tau_{U_0}^2$ use ($\sigma_e^2/n$) as correction factor: **True $\tau_{U_0}^2$ = Observed $\tau_{U_0}^2$ − ($\sigma_e^2/n$)**
- For example: observed level-2 $\tau_{U_0}^2$=4.65, level-1 $\sigma_e^2$=7.06, *n*=4
  - True $\tau_{U_0}^2$= 4.65 −(**7.60**/4) = **2.88** in empty means model
  - Add fixed linear time slope → reduce $\sigma_e^2$ from 7.06 to 2.17 ($R^2$ = .69)
  - But now True $\tau_{U_0}^2$= 4.65 −(**2.17**/4) = **4.10** in fixed linear time model

# Quantification of Random Effects Variances

- We can test if a random effect variance is significant, but the variance estimates are not likely to have inherent meaning
  - e.g., "I have a significant fixed linear time effect of $\gamma_{10}$ = **1.72**, so people increase by 1.72/time on average. I also have a significant random linear time slope variance of $\tau^2_{U_1}$ = **0.91**, so people need their own slopes (people change differently). But how much is a variance of **0.91**, really?"
- **95% Random Effects Confidence Intervals** can tell you
  - Can be calculated for each effect that is random in your model
  - Provide range around the fixed effect within which 95% of your sample is predicted to fall, based on your random effect variance:

$$\text{Random Effect 95\% CI} = \text{fixed effect} \pm \left(1.96 * \sqrt{\text{Random Variance}}\right)$$

$$\text{Linear Time Slope 95\% CI} = \gamma_{10} \pm \left(1.96 * \sqrt{\tau^2_{U_1}}\right) \rightarrow 1.72 \pm \left(1.96 * \sqrt{0.91}\right) = -0.15 \text{ to } 3.59$$

  - So although people improve on average, individual slopes are predicted to range from −0.15 to 3.59 (so some people may actually decline)



# Describing Within-Person Change in Longitudinal Data

- Topics:
  - Multilevel modeling notation and terminology
  - Fixed and random effects of linear time
  - Predicted variances and covariances from random slopes
  - Dependency and effect size in random effects models
  - **Describing nonlinear change: polynomial, piecewise models, and truly nonlinear models**
  - Fun with likelihood estimation and model comparisons

# Summary: Modeling Means and Variances

- We have two tasks in describing within-person change:

- **Choose a Model for the Means**
  - What kind of change in the outcome do we have **on average**?
  - What kind and how many **fixed effects** do we need to predict that mean change as parsimoniously but accurately as possible?

- **Choose a Model for the Variances**
  - What pattern do the variances and covariances of the outcome show over time because of **individual differences** in change?
  - What kind and how many **random effects** do we need to predict that pattern as parsimoniously but accurately as possible?



# Name that trajectory… Polynomial?

- Predict **mean change** with **polynomial fixed effects of time**:
  - Linear → *constant* amount of change (up or down)
  - Quadratic → *change* in linear rate of change (acceleration/deceleration)
  - Cubic → *change* in acceleration/deceleration of linear rate of change (known in physics as jerk, surge, or jolt)
  - Terms work together to describe curved trajectories

- **Can have polynomial fixed time slopes UP TO: *n* – 1***
  - 3 occasions = 2nd order ($time^2$)= Fixed Quadratic Time or less
  - 4 occasions = 3rd order ($time^3$) = Fixed Cubic Time or less

- Interpretable polynomials past cubic are rarely seen in practice

*$n-1$ rule can be broken in unbalanced data (but cautiously)

# Interpreting Quadratic Fixed Effects

**A Quadratic time effect is a two-way interaction: time*time**

- Fixed quadratic time = "**half the rate of acceleration/deceleration**"
- So to interpret it as how the linear time effect changes per unit time, **you must multiply the quadratic coefficient by 2**
- If fixed linear time slope = 4 at time 0, with quadratic slope = 0.3?
  - Instantaneous linear rate of Δ at time 0 = 4.0, at time 1 = 4.6…
- The "twice" part comes from taking the derivatives of the function:

$$\text{Intercept (Position) at Time T: } \hat{y}_T = 50.0 + 4.0T + 0.3T^2$$

$$\text{First Derivative (Velocity) at Time T: } \frac{d\hat{y}_T}{d(T)} = 4.0 + 0.6T$$

$$\text{Second Derivative (Acceleration) at Time T: } \frac{d^2\hat{y}_T}{d(T)} = 0.6$$



# Interpreting Quadratic Fixed Effects

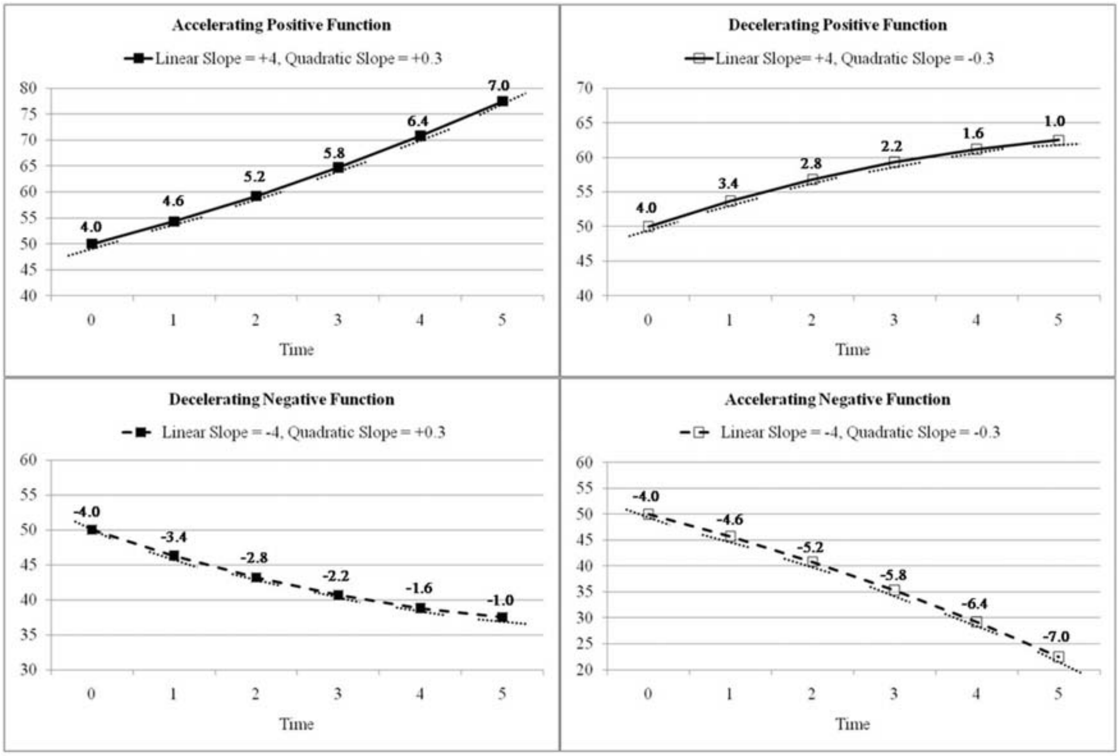
**A Quadratic time effect is a two-way interaction: time*time**

- Fixed quadratic time = "**half the rate of acceleration/deceleration**"
- So to interpret it as how the linear time effect changes per unit time, **you must multiply the quadratic coefficient by 2**
- If fixed linear time slope = 4 at time 0, with quadratic slope = 0.3?
  - Instantaneous linear rate of Δ at time 0 = 4.0, at time 1 = 4.6…
- The "twice" part also comes from what you remember about the role of interactions with respect to their constituent main effects:

$$\hat{y} = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ$$
$$\text{Effect of } X = \beta_1 + \beta_3 Z$$
$$\text{Effect of } Z = \beta_2 + \beta_3 X$$
$$\hat{y}_T = \beta_0 + \beta_1 \text{Time}_T + \_\_\_\_ + \beta_3 \text{Time}_T^2$$
$$\text{Effect of Time}_T = \beta_1 + 2\beta_3 \text{Time}_T$$

- Because time is interacting with itself, there is no second main effect in the model for the interaction to modify as usual. So the quadratic time effect gets applied <u>twice</u> to the <u>one</u> (main) linear effect of time.

# Examples of Fixed Quadratic Time Effects






# Conditionality of Polynomial Fixed Time Effects

- We've seen how main effects become conditional simple effects once they are part of an interaction
- The same is true for polynomial **fixed effects of time**:
  - ***Fixed Intercept Only?***
    - Fixed Intercept = predicted mean of Y *for any occasion (= grand mean)*
  - ***Add Fixed Linear Time?***
    - Fixed Intercept = **now** predicted mean of Y from linear time *at time=0* *(would be different if time was centered elsewhere)*
    - Fixed Linear Time = mean linear rate of change *across all occasions* *(would be the same if time was centered elsewhere)*
  - ***Add Fixed Quadratic Time?***
    - Fixed Intercept = still predicted mean of Y *at time=0 (but from quadratic model)* (*would be different if time was centered elsewhere*)
    - Fixed Linear Time = **now** mean linear rate of change *at time=0* *(would be different if time was centered elsewhere)*
    - Fixed Quadratic Time = half the mean rate of acceleration or deceleration of change *across all occasions* (i.e., the linear slope changes the same over time)

# Polynomial Fixed vs. Random Time Effects

- **Polynomial fixed effects** combine to describe mean trajectory over time (can have fixed slopes up to ***n* – 1**):
  - Fixed Intercept = Predicted mean level (at time 0)
  - Fixed Linear Time = Mean linear rate of change (at time 0)
  - Fixed Quadratic Time = Half of mean acceleration/deceleration in linear rate of change (2*quad is how the linear time slope changes per unit time if quadratic is highest order fixed effect of time)
- **Polynomial random effects** (individual deviations from the fixed effect) describe individual differences in those change parameters (can have random slopes up to ***n* – 2**):
  - Random Intercept = BP variance in level (at time 0)
  - Random Linear Time = BP variance in linear time slope (at time 0)
  - Random Quadratic Time = BP variance in half the rate of acceleration/deceleration of linear time slope (across all time if quadratic is highest-order random effect of time)



# Random Quadratic Time Model

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}Time_{ti} + \beta_{2i}Time_{ti}^2 + e_{ti}$

Level 2 Equations (one per β):

$\beta_{0i} = \gamma_{00} + U_{0i}$

- $\beta_{0i}$: **Intercept** for person *i*
- $\gamma_{00}$: Fixed (mean) **Intercept**
- $U_{0i}$: Random (Deviation) **Intercept**

$\beta_{1i} = \gamma_{10} + U_{1i}$

- $\beta_{1i}$: **Linear Slope** for person *i*
- $\gamma_{10}$: Fixed (mean) **Linear Slope**
- $U_{1i}$: Random (Deviation) **Linear Slope**

$\beta_{2i} = \gamma_{20} + U_{2i}$

- $\beta_{2i}$: **Quad Slope** for person *i*
- $\gamma_{20}$: Fixed (mean) **Quad Slope**
- $U_{2i}$: Random (Deviation) **Quad Slope**

**Fixed Effect Subscripts:**
1st = which Level 1 term
2nd = which Level 2 term

**Number of Possible Slopes by Number of Occasions (*n*):**

# Fixed slopes = $n - 1$
# Random slopes = $n - 2$

Need $n = 4$ occasions to fit random quadratic time model

# Conditionality of Polynomial Random Effects

- We saw previously that lower-order fixed effects of time are conditional on higher-order polynomial fixed effects of time
- The same is true for polynomial **random effects of time**:
  - ***Random Intercept Only?***
    - Random Intercept = BP variance *for any occasion* in predicted mean Y (= variance in grand mean because individual lines are parallel)
  - ***Add Random Linear Time?***
    - Random Intercept = **now** BP variance *at time=0* in predicted mean Y *(would be different if time was centered elsewhere)*
    - Random Linear Time = BP variance *across all occasions* in linear rate of change *(would be the same if time was centered elsewhere)*
  - ***Add Random Quadratic Time?***
    - Random Intercept = still BP variance *at time=0* in predicted mean Y
    - Random Linear Time = **now** BP variance *at time=0* in linear rate of change *(would be different if time was centered elsewhere)*
    - Random Quadratic Time = BP variance *across all occasions* in half of accel/decel of change *(would be the same if time was centered elsewhere)*



# Random Effects Allowed by #Occasions

| | Data | G Matrix | R Matrix | Variance Model # Parameters |
|---|---|---|---|---|
| *n*=2 occasions<br>**3** unique pieces of information | $\begin{bmatrix} \sigma_1^2 & \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$ | $\begin{bmatrix} \tau_{U_0}^2 \end{bmatrix}$<br>**Random Intercept only** | $\begin{bmatrix} \sigma_e^2 & 0 \\ 0 & \sigma_e^2 \end{bmatrix}$ | **2** |
| *n*=3 occasions<br>**6** unique pieces of information | $\begin{bmatrix} \sigma_1^2 & & \\ \sigma_{21} & \sigma_2^2 & \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$ | $\begin{bmatrix} \tau_{U_0}^2 & \\ \tau_{U_{01}} & \tau_{U_1}^2 \end{bmatrix}$<br>**Up to 1 Random slope** | $\begin{bmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_e^2 & 0 \\ 0 & 0 & \sigma_e^2 \end{bmatrix}$ | **4** |
| *n*=4 occasions<br>**10** unique pieces of information | $\begin{bmatrix} \sigma_1^2 & & & \\ \sigma_{21} & \sigma_2^2 & & \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$ | $\begin{bmatrix} \tau_{U_0}^2 & & \\ \tau_{U_{01}} & \tau_{U_1}^2 & \\ \tau_{U_{02}} & \tau_{U_{12}} & \tau_{U_2}^2 \end{bmatrix}$<br>**Up to 2 Random slopes** | $\begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & \sigma_e^2 \end{bmatrix}$ | **7** |

# Predicted **V** Matrix from Polynomial Random Effects Models

- **<u>Random linear model?</u>** Variance has a **quadratic** dependence on time
  - Variance will be at a minimum when time = $-Cov(U_0,U_1)/Var(U_1)$, and will increase parabolically and symmetrically over time
  - **Predicted variance** at each occasion and covariance between A and B:

    $Var(y_{time}) = Var(e_t) + Var(U_0) + 2Cov(U_0,U_1)(\mathbf{time_t}) + Var(U_1)(\mathbf{time_t^2})$

    $Cov(y_A,y_B) = Var(U_0) + Cov(U_0,U_1)(A + B) + Var(U_1)(AB)$

- **<u>Random quadratic model?</u>** Variance has a **quartic** dependence on time

  $Var(y_{time}) = Var(e_t) + Var(U_0) + 2Cov(U_0,U_1)(\mathbf{time_t}) + Var(U_1)(\mathbf{time_t^2}) + 2Cov(U_0,U_2)(\mathbf{time_t^2}) + 2Cov(U_1,U_2)(\mathbf{time_t^3}) + Var(U_2)(\mathbf{time_t^4})$

  $Cov(y_A,y_B) = Var(U_0) + Cov(U_0,U_1)(A + B) + Var(U_1)(AB) + Cov(U_0,U_2)(A^2 + B^2) + Cov(U_1,U_2)[(AB^2)+(A^2B)] + Var(U_2)(A^2B^2)$

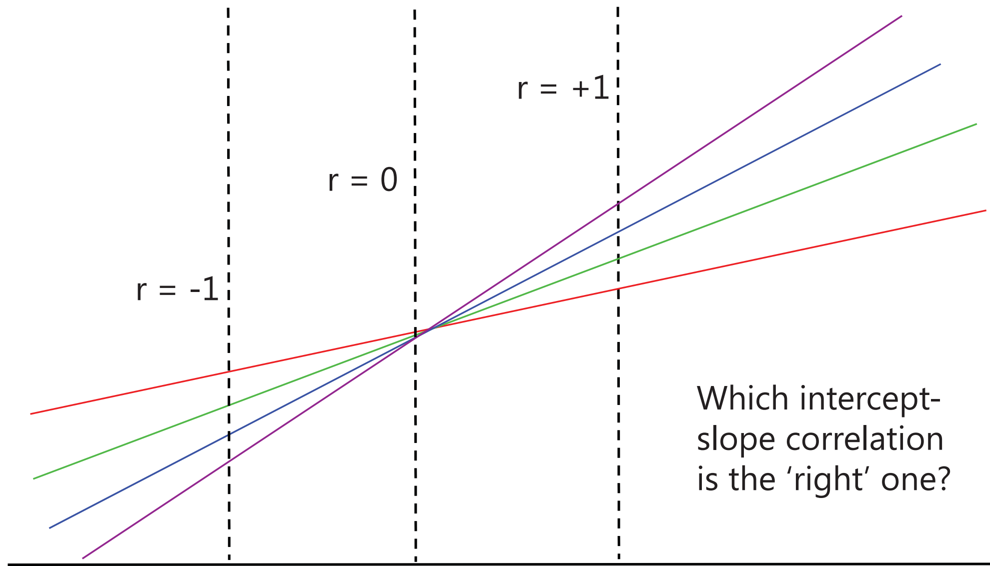
- *The point of the story: random effects of time are a way of allowing the variances and covariances to differ over time in specific, time-dependent patterns (that result from differential individual change over time).*



# Rules for Polynomial Models (and in general for fixed and random effects)

- On the same side of the model (means or variances side), lower-order effects stay in EVEN IF NONSIGNIFICANT (for correct interpretation)
  - e.g., Significant *fixed* quadratic? Keep the *fixed* linear
  - e.g., Significant *random* quadratic? Keep the *random* linear
- Also remember—you can have a significant random effect EVEN IF the corresponding fixed effect is not significant (keep it anyway):
  - e.g., Fixed linear not significant, but random linear is significant?
    → No linear change *on average*, but significant individual differences in change
- Language: A random effect supersedes a fixed effect:
  - If <u>Fixed</u> = intercept, linear, quad; <u>Random</u> = intercept, linear, quad?
    - Call it a "Random quadratic model" (implies everything beneath those terms)
  - If <u>Fixed</u> = intercept, linear, quad; <u>Random</u> = intercept, linear?
    - Call it a "Fixed quadratic, random linear model" (distinguishes no random quad)
- Intercept-slope correlation depends largely on centering of time…

# Correlation between Random Intercept and Random Linear Slope depends on time 0




Which intercept-slope correlation is the 'right' one?

!! Nonparallel lines will eventually cross.



# Correlations among polynomial slopes

Session Centered at 1:

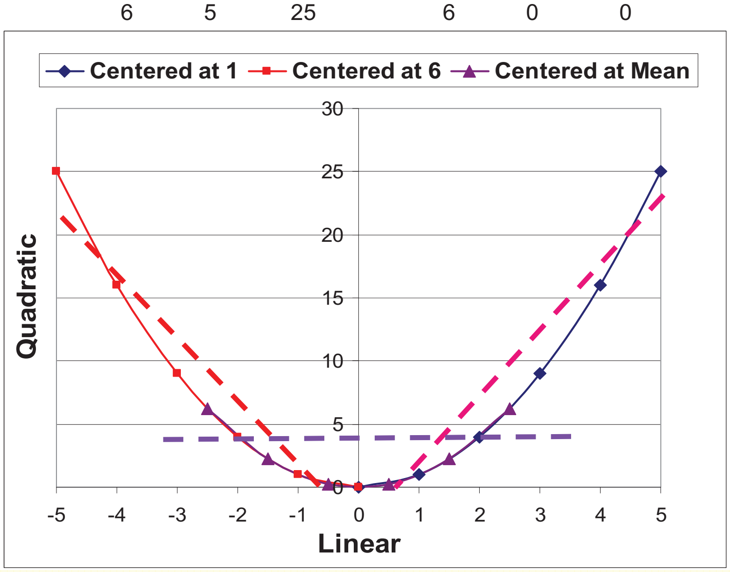
| Session | Linear | Quadratic |
|---|---|---|
| 1 | 0 | 0 |
| 2 | 1 | 1 |
| 3 | 2 | 4 |
| 4 | 3 | 9 |
| 5 | 4 | 16 |
| 6 | 5 | 25 |

Session Centered at 6:

| Session | Linear | Quadratic |
|---|---|---|
| 1 | -5 | 25 |
| 2 | -4 | 16 |
| 3 | -3 | 9 |
| 4 | -2 | 4 |
| 5 | -1 | 1 |
| 6 | 0 | 0 |

Session Centered at Mean:

| Session | Linear | Quadratic |
|---|---|---|
| 1 | -2.5 | 6.25 |
| 2 | -1.5 | 2.25 |
| 3 | -0.5 | 0.25 |
| 4 | 0.5 | 0.25 |
| 5 | 1.5 | 2.25 |
| 6 | 2.5 | 6.25 |

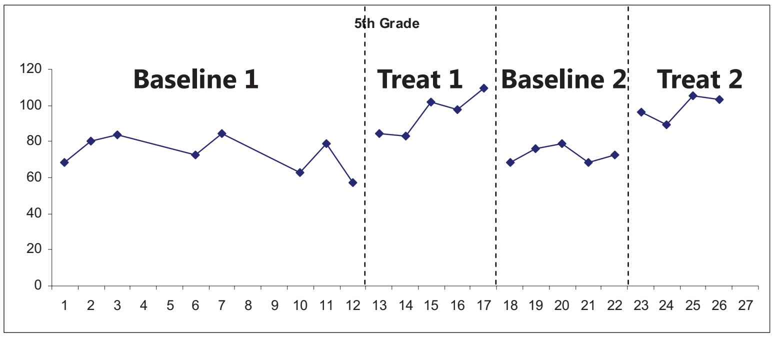



Correlations among polynomial effects of time can be induced by centering time near the start or near the end.

Therefore, these correlations will be *most* interpretable when centering time at its mean instead.

# Other Random Effects Models of Change

- **Piecewise models**: Discrete slopes for discrete phases of time
  - Separate terms describe sections of overall trajectories
  - Useful for examining change in intercepts and slopes before/after discrete events (changes in policy, interventions)
  - **Must know where the break point is ahead of time!**




**Piecewise Model:**

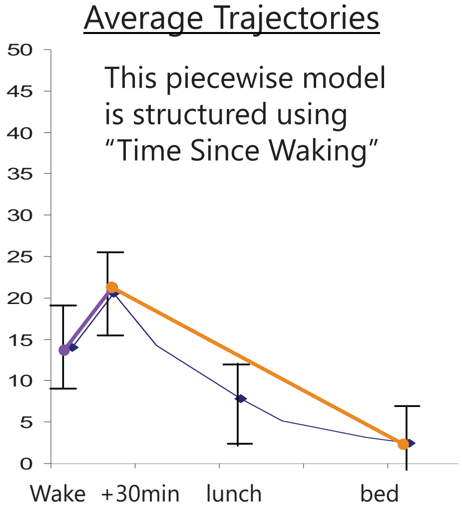
4 slopes (one per phase)

3 "jumps" (shift in intercept between phases)



# Example of Daily Cortisol Fluctuation: Morning Rise and Afternoon Decline

Average Trajectories




**SAS Code** to create two piecewise slopes from continuous time of day in stacked data:

```
IF occasion=1 THEN DO;
    P1=0;              P2=0; END;
IF occasion=2 THEN DO;
    P1= time2-time1; P2=0; END;
IF occasion=3 THEN DO;
    P1= time2-time1; P2=time3-time2; END;
IF occasion=4 THEN DO;
    P1= time2-time1; P2=time4-time2; END;
```

Note that a quadratic slope may be necessary for the afternoon decline slope!

# Random Two-Slope Piecewise Model

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}\text{Slope1}_{ti} + \beta_{2i}\text{Slope2}_{ti} + e_{ti}$

Level 2 Equations (one per β):

$\beta_{0i} = \gamma_{00} + U_{0i}$

**Intercept** for person *i* = Fixed (mean) **Intercept** + Random (Deviation) **Intercept**

$\beta_{1i} = \gamma_{10} + U_{1i}$

**Slope1** for person *i* = Fixed (mean) **Slope1** + Random (Deviation) **Slope1**

$\beta_{2i} = \gamma_{20} + U_{2i}$

**Slope2** for person *i* = Fixed (mean) **Slope2** + Random (Deviation) **Slope2**

**Fixed Effect Subscripts:**

1st = which Level 1 term
2nd = which Level 2 term

**Number of Possible Slopes by Number of Occasions ($n$):**

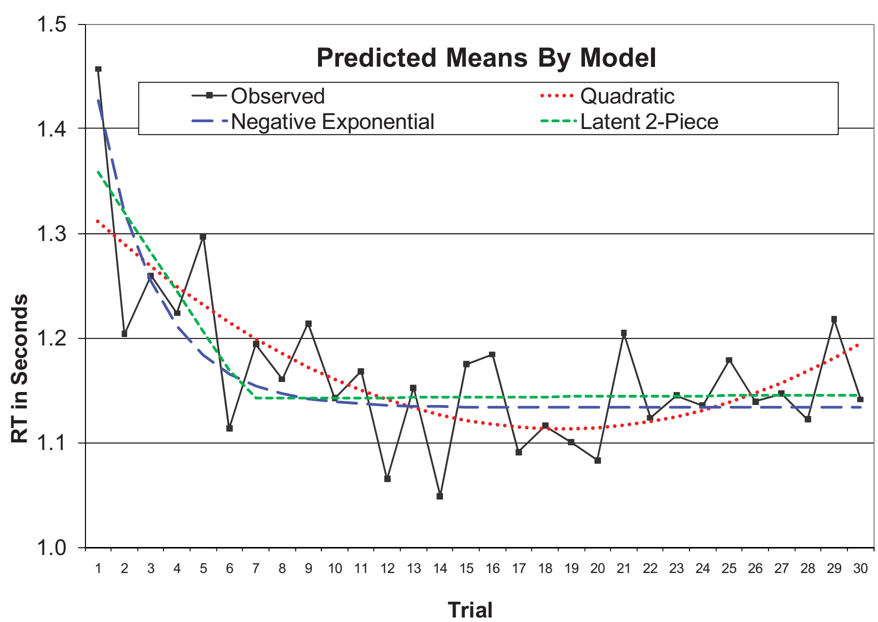
# Fixed slopes = $n - 1$
# Random slopes = $n - 2$

Need $n$ = 4 occasions to fit random two-slope model

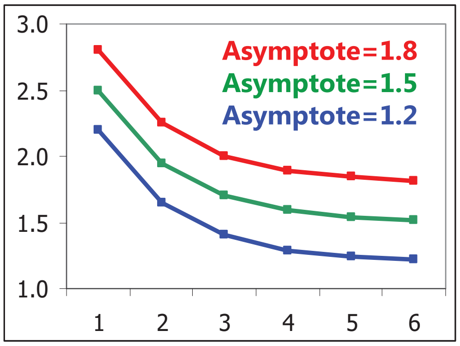


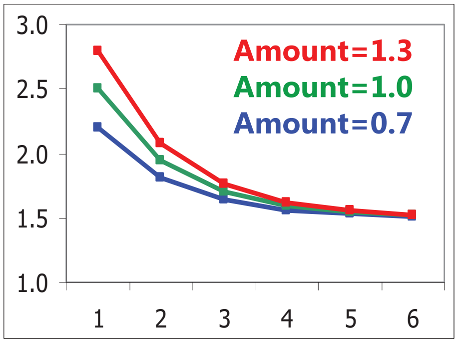
# Other Random Effects for Change

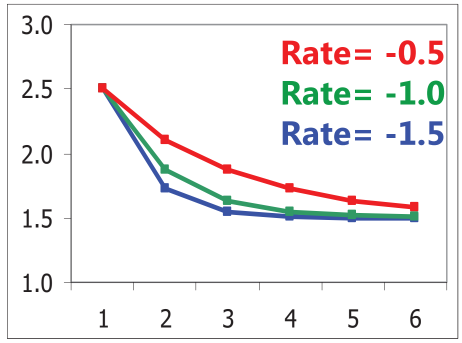
- **Truly nonlinear models**: Non-additive terms to describe change
  - Models can include **asymptotes** (so change can "shut off" as needed)
  - Include **power** and **exponential** functions (see chapter 6 for references)



# (Negative) Exponential Model Parameters

1) 


2) 


3) 


1) Different **Asymptotes**, same amount and rate

2) Different **Amounts**, same asymptote and rate

3) Different **Rates**, same asymptote and amount



# Exponential Model (3 Random Effects)

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}*\exp(\beta_{2i}*Time_{ti}) + e_{ti}$

Level 2 Equations (one per β):

$\beta_{0i} = \gamma_{00} + U_{0i}$

- $\beta_{0i}$: **Asymptote** for person *i*
- $\gamma_{00}$: Fixed (mean) **Asymptote**
- $U_{0i}$: Random (Deviation) **Asymptote**

$\beta_{1i} = \gamma_{10} + U_{1i}$

- $\beta_{1i}$: **Amount** for person *i*
- $\gamma_{10}$: Fixed (mean) **Amount**
- $U_{1i}$: Random (Deviation) **Amount**

$\beta_{2i} = \gamma_{20} + U_{2i}$

- $\beta_{2i}$: **Rate** for person *i*
- $\gamma_{20}$: Fixed (mean) **Rate**
- $U_{2i}$: Random (Deviation) **Rate**

**Fixed Effect Subscripts:**
1st = which Level 1 term
2nd = which Level 2 term

**Number of Possible Slopes by Number of Occasions (*n*):**

\# Fixed slopes = $n - 1$
\# Random slopes = $n - 2$

Also need 4 occasions to fit random exponential model

(Likely need way more occasions to find $U_{2i}$, though)

# Describing Within-Person Change in Longitudinal Data

- Topics:
  - Multilevel modeling notation and terminology
  - Fixed and random effects of linear time
  - Predicted variances and covariances from random slopes
  - Dependency and effect size in random effects models
  - Describing nonlinear change: polynomial, piecewise models, and truly nonlinear models
  - **Fun with likelihood estimation and model comparisons**



# 3 Decision Points for Model Comparisons

1. Are the models **nested** or **non-nested**?
   - Nested: have to add OR subtract effects to go from one to other
     - Can conduct significance tests for improvement in fit
   - Non-nested: have to add AND subtract effects
     - No significance tests available for these comparisons

2. Differ in model for the **means**, **variances**, or **both**?
   - Means? Can only use ML $-2\Delta LL$ tests (or *p*-value of each fixed effect)
   - Variances? Can use ML (or preferably REML) $-2\Delta LL$ tests, no *p*-values
   - Both sides? Can only use ML $-2\Delta LL$ tests

3. Models estimated using **ML** or **REML**?
   - ML: All model comparisons are ok
   - REML: Model comparisons are ok for the variance parameters only

# Likelihood-Based Model Comparisons

- Relative model fit is indexed by a "**deviance**" statistic → **−2LL**
  - Log of likelihood (**LL = total data height**) of observing the data given model parameters, −2*LL so that the differences between model LL values follow $\sim\chi^2$
  - **−2LL is a measure of BADNESS of fit, so smaller values = better models**
  - Models are compared using their deviance values (significance tests)
  - Two estimation flavors (labeled as −2 log likelihood in SAS, SPSS, but given as LL instead in STATA): Maximum Likelihood (**ML**) or Restricted (Residual) ML (**REML**)

- Fit is also indexed by **Information Criteria** that reflect **−2LL** deviance AND # parameters used and/or sample size
  - **AIC** = Akaike IC = **−2LL** + 2 *(#parameters)
  - **BIC** = Bayesian IC = **−2LL** + log(N)*(#parameters) → penalty for complexity
  - In ML → #parameters = all parameters (means and variances models)
  - In REML → #parameters = variance model parameters only (except in STATA!)
  - No significance tests or critical values, just "smaller is better"



# −2ΔLL (i.e., LRT, Deviance) Tests: (models must use the same estimator & *N*)

1. Calculate −2ΔLL: $(-2LL_{fewer}) - (-2LL_{more})$
2. Calculate Δdf: $(\#\ Parms_{more}) - (\#\ Parms_{fewer})$
3. Compare −2ΔLL to $\chi^2$ distribution with df = Δdf
   CHIDIST function in excel will give exact p-values for the difference test; so will STATA

1. & 2. must be positive values!

- Fixed effects $p<.05$: −2ΔLL(1)>3.84, −2ΔLL(2)>5.99, −2ΔLL(3)>7.82
- Some controversy about −2ΔLL tests when testing random effects variances that cannot be negative (i.e., the "boundary problem")
  - $\chi^2$ is not distributed as usual (mean=df) → is actually a mixture $\chi^2$ with df and df−1, so using the critical $\chi^2$ for actual df results in conservative model comparison test
  - e.g., −2ΔLL(df=2)>5.99, whereas −2ΔLL(df=mixture of 1,2)>5.14
- Two proposed solutions when testing random effects variances:
  - For random intercepts, can use a 1-tailed test ($\chi^2$ for $p < .10$): −2ΔLL(1)>2.71
  - Use mixture $p$-value = 0.5*prob($\chi^2_{df-1} > -2\Delta LL$) + 0.5*prob($\chi^2_{df} > -2\Delta LL$)
  - In practice these assume no relationship among how well variance parameters are estimated, which is suspect → I tend to just use the conservative test and call it good

# Critical Values for 50:50 $\chi^2$ Mixtures

| | Significance Level | | | | |
|---|---|---|---|---|---|
| **df (q)** | **0.10** | **0.05** | **0.025** | **0.01** | **0.005** |
| **0 vs. 1** | 1.64 | 2.71 | 3.84 | 5.41 | 6.63 |
| **1 vs. 2** | 3.81 | 5.14 | 6.48 | 8.27 | 9.63 |
| **2 vs. 3** | 5.53 | 7.05 | 8.54 | 10.50 | 11.97 |
| **3 vs. 4** | 7.09 | 8.76 | 10.38 | 12.48 | 14.04 |
| **4 vs. 5** | 8.57 | 10.37 | 12.10 | 14.32 | 15.97 |
| **5 vs. 6** | 10.00 | 11.91 | 13.74 | 16.07 | 17.79 |
| **6 vs. 7** | 11.38 | 13.40 | 15.32 | 17.76 | 19.54 |
| **7 vs. 8** | 12.74 | 14.85 | 16.86 | 19.38 | 21.23 |
| **8 vs. 9** | 14.07 | 16.27 | 18.35 | 20.97 | 22.88 |
| **9 vs. 10** | 15.38 | 17.67 | 19.82 | 22.52 | 24.49 |
| **10 vs. 11** | 16.67 | 19.04 | 21.27 | 24.05 | 26.07 |

This may work ok if only one new parameter is bounded … for example:

+ Random Intercept
df=1: 2.71 vs. 3.84

+ Random Linear
df=2: 5.14 vs. 5.99

+ Random Quad
df=3: 7.05 vs. 7.82

Critical values such that the right-hand tail probability =
$0.5 \times \Pr(\chi^2_q > c) + 0.5 \times \Pr(\chi^2_{q+1} > c)$

Source: Appendix C (p. 484) from Fitzmaurice, Laird, & Ware (2004). *Applied Longitudinal Analysis*. Hoboken, NJ: Wiley



# ML vs. REML

Remember "population" vs. "sample" formulas for calculating variance?

Population: $\sigma_e^2 = \dfrac{\sum_{i=1}^{N}(y_i - \mu)^2}{N}$ Sample: $\sigma_e^2 = \dfrac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1}$

| **All comparisons must have same N!!!** | **ML** | **REML** |
|---|---|---|
| To select, type… | METHOD=ML<br>(-2 log likelihood) | METHOD=REML *default*<br>(-2 res log likelihood) |
| In estimating variances, it treats fixed effects as… | **Known** (df for having to also estimate fixed effects is not factored in) | **Unknown** (df for having to estimate fixed effects is factored in) |
| So, in small samples, L2 variances will be… | **Too small** (less difference after N=30-50 or so) | **Unbiased** (correct) |
| But because it indexes the fit of the… | **Entire model**<br>(means + variances) | **Variances model only** |
| You can compare models differing in… | **Fixed and/or random effects** (either/both) | **Random effects only**<br>(same fixed effects) |

# Rules for Comparing Multilevel Models

**All observations must be the same across models!**

**Compare Models Differing In:**

| **Type of Comparison:** | **Means Model (Fixed) Only** | **Variance Model (Random) Only** | **Both Means and Variances Model (Fixed and Random)** |
|---|---|---|---|
| **Nested?** YES, can do significance tests via… | Fixed effect *p*-values from ML or REML *-- OR --* ML −2ΔLL only (NO REML −2ΔLL) | NO *p*-values<br>REML −2ΔLL (ML −2ΔLL is ok if big N) | ML −2ΔLL only (NO REML −2ΔLL) |
| **Non-Nested?** NO signif. tests, instead see… | ML AIC, BIC (NO REML AIC, BIC) | REML AIC, BIC (ML ok if big N) | ML AIC, BIC only (NO REML AIC, BIC) |

Nested = one model is a direct subset of the other

Non-Nested = one model is not a direct subset of the other



# Summary: Model Comparisons

- Significance of **fixed effects** can be tested with EITHER their ***p*-values** OR **ML** −2ΔLL (LRT, deviance difference) tests
  - *p*-value → Is EACH of these effects significant? (fine under ML or REML)
  - ML −2ΔLL test → Does this SET of predictors make my model better?
  - *REML −2ΔLL tests are WRONG for comparing models differing in fixed effects*
- Significance of **random effects** can only be tested with **−2ΔLL tests** (preferably using REML; here ML is not wrong, but results in too small variance components and fixed effect SEs in smaller samples)
  - Can get *p*-values as part of output but *shouldn't* use them
  - #parms added (df) should always include the random effect covariances
- My recommended approach to building models:
  - Stay in REML (for best estimates), test new fixed effects with their *p*-values
  - THEN add new random effects, testing −2ΔLL against previous model

# Example Sequence for Testing Fixed and Random Polynomial Effects of Time

Build up fixed and random effects simultaneously:

1. Empty Means, Random Intercept → to calculate ICC
2. Fixed Linear, Random Intercept → check fixed linear *p*-value
3. Random Linear → check −2ΔLL(df≈2) for random linear variance
4. Fixed Quadratic, Random Linear → check fixed quadratic *p*-value
5. Random Quadratic → check −2ΔLL(df≈3) for random quadratic variance
6. …….

*** In general: Can use **REML** for all models, so long as you:

→ Test significance of new **fixed** effects by their ***p*-values**

→ Test significance of new **random** effects in separate step by **−2ΔLL**

→ Also see if AIC and BIC are smaller when adding random effects



# Time-Invariant Predictors in Longitudinal Models

- Topics:
  - **Missing predictors in MLM**
  - Effects of time-invariant predictors
  - Fixed, systematically varying, and random level-1 effects
  - Model building strategies and assessing significance

# Summary of Steps in Unconditional Longitudinal Modeling

**For all outcomes:**

1. Empty Model; Calculate ICC
2. Decide on a metric of time
3. Decide on a centering point
4. Estimate means model and plot individual trajectories

**If your outcome shows systematic change:**

5. Evaluate fixed and random effects of time
6. Still consider possible alternative models for the residuals (**R** matrix)

**If your outcome does NOT show ANY systematic change:**

5. Evaluate alternative models for the variances (**G**+**R**, or **R**)



# Random Effects Models for the Variance

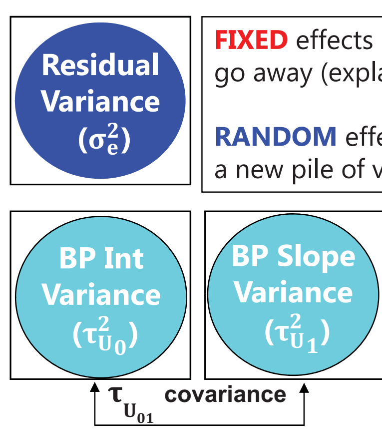
- Each source of correlation or dependency goes into a new variance component (or pile of variance) until each source meets the usual assumptions of GLM: normality, independence, constant variance
- **Example 2-level longitudinal model:**

**Level 1 (one source of) <u>Within-Person Variation</u>: gets accounted for by time-level predictors**

**FIXED** effects make variance go away (explain variance).

**RANDOM** effects just make a new pile of variance.

**Level 2 (two sources of) <u>Between-Person Variation</u>: gets accounted for by person-level predictors**




**Now we get to add predictors to account for each pile!**

# Missing Data in MLM Software

- Common misconceptions about how MLM "handles" missing data
- Most MLM programs analyze only COMPLETE CASES
  - Does NOT require listwise deletion of *whole persons*
  - DOES delete any incomplete cases (occasions within a person)
- Observations missing predictors OR outcomes are not included!
  - **Time** is (probably) measured for **everyone**
  - **Predictors may NOT be measured for everyone**
  - *N* may change due to missing data for different predictors across models
- You may need to think about what predictors you want to examine PRIOR to model building, and pre-select your sample accordingly
  - Models and model fit statistics −2LL, AIC, and BIC are only **directly comparable** if they include the **exact same observations (LL is sum of each height)**
  - Will have less statistical power as a result of removing incomplete cases



# Be Careful of Missing Predictors!

**Multivariate (wide) data → stacked (long) data**

| ID | T1 | T2 | T3 | T4 | Person Pred | T1 Pred | T2 Pred | T3 Pred | T4 Pred |
|---|---|---|---|---|---|---|---|---|---|
| **100** | 5 | 6 | 8 | 12 | 50 | 4 | 6 | 7 | . |
| **101** | 4 | 7 | . | 11 | . | 7 | . | 4 | 9 |

| Row | ID | Time | DV | Person Pred | Time Pred |
|---|---|---|---|---|---|
| **1** | **100** | 1 | 5 | 50 | 4 |
| **2** | **100** | 2 | 6 | 50 | 6 |
| **3** | **100** | 3 | 8 | 50 | 7 |
| **4** | **100** | 4 | 12 | 50 | . |
| **5** | **101** | 1 | 4 | . | 7 |
| **6** | **101** | 2 | 7 | . | . |
| **7** | **101** | 3 | . | . | 4 |
| **8** | **101** | 4 | 11 | . | 9 |

**Only rows with complete data get used – for each model, which rows get used in MIXED?**

| | |
|---|---|
| Model with Time → DV: | 1-6, 8 |
| Model with Time, Time Pred → DV: | 1-3, 5, 8 |
| Model with Time, Person Pred → DV: | 1-4 |
| Model with Time, Time Pred, & Person Pred → DV: | 1-3 |

## So what does this mean for missing data in MLM?

- **Missing outcomes are assumed MAR**
  - Because the likelihood function is for predicted Y, just estimated on whatever Y responses a person does have (can be incomplete)
- **Missing time-varying predictors are MAR-to-MCAR ish**
  - Would be MCAR because X is not in the likelihood function (is Y given X instead), but other occasions may have predictors (so MAR-ish)
- **Missing time-invariant predictors are assumed MCAR**
  - Because the predictor would be missing for all occasions, whole people will be deleted (may lead to bias)

- Missingness on predictors can be accommodated:
  - In Multilevel SEM with certain assumptions (≈ outcomes then)
  - Via multilevel multiple imputation in Mplus v 6.0+ (but careful!)
    - Must preserve all effects of potential interest in imputation model, including random effects; −2ΔLL tests are not done in same way



# Time-Invariant Predictors in Longitudinal Models

- Topics:
  - Missing predictors in MLM
  - **Effects of time-invariant predictors**
  - Fixed, systematically varying, and random level-1 effects
  - Model building strategies and assessing significance

# Modeling Time-Invariant Predictors

<u>What independent variables can be time-invariant predictors?</u>

- Also known as "person-level" or "level-2" predictors
- Include substantive predictors, controls, and predictors of missingness
- Can be anything that **does not change across time** (e.g., Biological Sex)
- Can be anything that **is not likely to change across the study,** but you may have to argue for this (e.g., Parenting Strategies, SES)
- Can be anything that **does change across the study**…
  - But you have **only measured once**
    - Limit conclusions to variable's status at time of measurement
    - e.g., "Parenting Strategies at age 10"
  - Or **is perfectly correlated with time** (age, time to event)
    - Would use Age at Baseline, or Time to Event *from Baseline* instead



# Centering Time-Invariant Predictors

- <u>Very useful to center all predictors such that 0 is a meaningful value:</u>
  - Same significance level of main effect, different interpretation of intercept
  - Different (more interpretable) main effects within higher-order interactions
    - With interactions, main effects = simple effects when other predictor = 0
- <u>Choices for centering **continuous** predictors:</u>
  - <u>At Mean</u>: Reference point is *average level of predictor within the sample*
    - Useful if predictor is on arbitrary metric (e.g., unfamiliar test)
  - <u>Better → At Meaningful Point</u>: Reference point is *chosen level of predictor*
    - Useful if predictor is already on a meaningful metric (e.g., age, education)
- <u>Choices for centering **categorical** predictors:</u>
  - Re-code group so that your chosen reference group = **reference (0) category!** (highest is the default in SAS and SPSS; lowest is default in STATA)
  - I do not recommend mean-centering categorical predictors (because who is at the mean of a categorical variable ?!?)

# Main Effects of Predictors within Interactions

- Main effects of predictors within interactions should remain in the model regardless of whether or not they are significant
  - An interaction is an over-additive (enhancing) or under-additive (dampening) effect, so *what it is additive to* must be included

- The role of a two-way interaction is to adjust its main effects…
- However, the idea of a "main effect" no longer applies… each main effect is ***conditional*** on the interacting predictor = 0

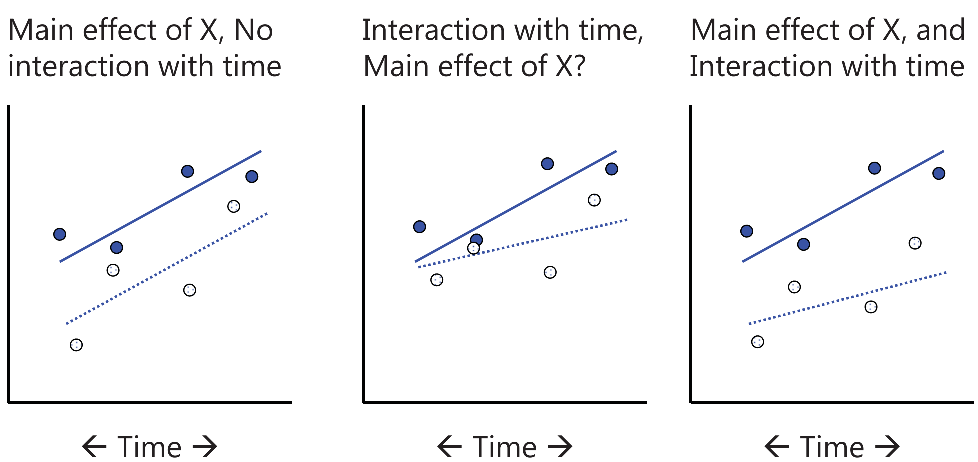
- e.g., Model of Y = W, X, Z, X*Z:
  - The effect of W is still a "main effect" because it is not part of an interaction
  - The effect of X is now the conditional main effect of X *specifically when Z=0*
  - The effect of Z is now the conditional main effect of Z *specifically when X=0*

- The trick is keeping track of what 0 means for every interacting predictor, which depends on the way each predictor is being represented, as determined by you, or by the software without you!



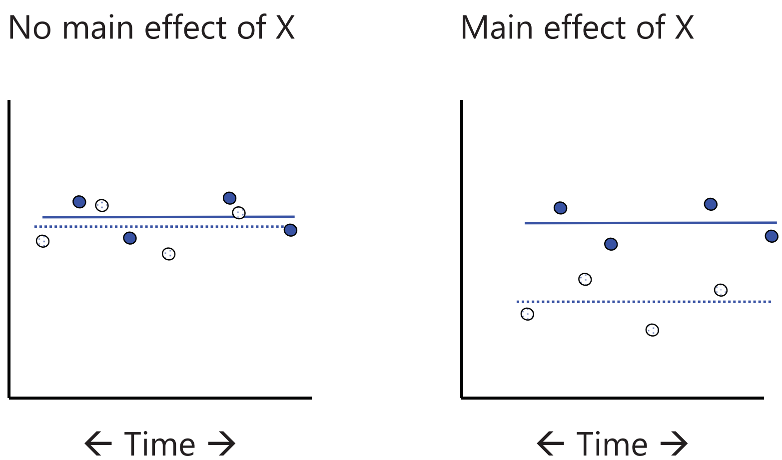
# The Role of Time-Invariant Predictors in the Model for the Means

- **In Within-Person Change Models** → Adjust growth curve



# The Role of Time-Invariant Predictors in the Model for the Means

- In **Within-Person Fluctuation Models** → Adjust mean level



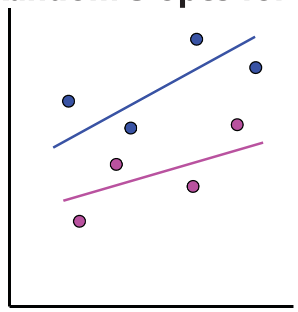



# The Role of Time-Invariant Predictors in the Model for the Variance

- In addition to fixed effects in the model for the means, time-invariant predictors can allow be used to allow **heterogeneity of variance** at their level or below
- e.g., Sex as a predictor of heterogeneity of variance:
  - **At level 2**: amount of individual differences in intercepts/slopes differs between boys and girls (i.e., one group is more variable)
  - **At level 1**: amount of within-person residual variation differs between boys and girls
    - In within-person **fluctuation** model: differential fluctuation over time
    - In within-person **change** model: differential fluctuation/variation remaining after controlling for fixed and random effects of time
- These models are harder to estimate and may require custom software (e.g., NLMIXED in SAS)

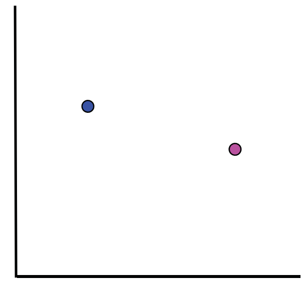
# Why Level-2 Predictors Cannot Have Random Effects in 2-Level Models

**Random Slopes for Time**



Time
(or Any Level-1 Predictor)

**Random Slopes for Sex?**



Sex
(or any Level-2 Predictor)

**You cannot make a line out of a dot, so level-2 effects cannot vary randomly over persons.**



# Education as a Time-Invariant Predictor: Example using a Random Quadratic Time Model

- Main Effect of Education = Education*Intercept Interaction
  - Moderates the intercept → Increase or decrease in expected outcome at time 0 for every year of education
- Effect of Education on Linear Time = Education*Time Interaction
  - Moderates the linear time slope → Increase or decrease in expected rate of change at time 0 for every year of education
- Effect of Education on Quadratic Time = Education*Time$^2$ Interaction
  - Moderates the quadratic time slope → Increase or decrease in half of expected acceleration/deceleration of linear rate of change for every year of education

# Education (12 years = 0) as a Time-Invariant Predictor: Example using a Random Quadratic Time Model

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}Time_{ti} + \beta_{2i}Time_{ti}^2 + e_{ti}$

Level 2 Equations (one per β):

$\beta_{0i} = \gamma_{00} + \gamma_{01}Ed_i + U_{0i}$

- $\beta_{0i}$: Intercept for person *i*
- $\gamma_{00}$: Fixed Intercept when Time=0 and Ed=12
- $\gamma_{01}$: Δ in Intercept per unit Δ in Ed
- $U_{0i}$: Random (Deviation) Intercept after controlling for Ed

$\beta_{1i} = \gamma_{10} + \gamma_{11}Ed_i + U_{1i}$

- $\beta_{1i}$: Linear Slope for person *i*
- $\gamma_{10}$: Fixed Linear Time Slope when Time=0 and Ed=12
- $\gamma_{11}$: Δ in Linear Time Slope per unit Δ in Ed (=Ed*time)
- $U_{1i}$: Random (Deviation) Linear Time Slope after controlling for Ed

$\beta_{2i} = \gamma_{20} + \gamma_{21}Ed_i + U_{2i}$

- $\beta_{2i}$: Quad Slope for person *i*
- $\gamma_{20}$: Fixed Quad Time Slope when Ed = 12
- $\gamma_{21}$: Δ in Quad Time Slope per unit Δ in Ed (=Ed*time$^2$)
- $U_{2i}$: Random (Deviation) Quad Time Slope after controlling for Ed



# Education (12 years = 0) as a Time-Invariant Predictor: Example using a Random Quadratic Time Model

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}Time_{ti} + \beta_{2i}Time_{ti}^2 + e_{ti}$

Level 2 Equations (one per β):

$\beta_{0i} = \gamma_{00} + \gamma_{01}Ed_i + U_{0i}$

$\beta_{1i} = \gamma_{10} + \gamma_{11}Ed_i + U_{1i}$

$\beta_{2i} = \gamma_{20} + \gamma_{21}Ed_i + U_{2i}$

$\gamma_{11}$ and $\gamma_{21}$ are known as "**cross-level**" interactions (level-1 predictor by level-2 predictor)

- Composite equation:
- $y_{ti} = (\gamma_{00} + \gamma_{01}Ed_i + U_{0i}) + (\gamma_{10} + \gamma_{11}Ed_i + U_{1i})Time_{ti} + (\gamma_{20} + \gamma_{21}Ed_i + U_{2i})Time_{ti}^2 + e_{ti}$

# Time-Invariant Predictors in Longitudinal Models

- Topics:
  - Missing predictors in MLM
  - Effects of time-invariant predictors
  - **Fixed, systematically varying, and random level-1 effects**
  - Model building strategies and assessing significance



# Fixed Effects of Time-Invariant Predictors

- Question of interest: Why do people change differently?
  - We're trying to predict individual differences in intercepts and slopes (i.e., reduce level-2 random effects variances)
  - So level-2 random effects variances become 'conditional' on predictors → actually random effects variances *left over*

$$\beta_{0i} = \gamma_{00} + U_{0i} \qquad \beta_{0i} = \gamma_{00} + \gamma_{01}Ed_i + U_{0i}$$
$$\beta_{1i} = \gamma_{10} + U_{1i} \longrightarrow \beta_{1i} = \gamma_{10} + \gamma_{11}Ed_i + U_{1i}$$
$$\beta_{2i} = \gamma_{20} + U_{2i} \qquad \beta_{2i} = \gamma_{20} + \gamma_{21}Ed_i + U_{2i}$$

  - Can calculate pseudo-$R^2$ for each level-2 random effect variance between models with *fewer* versus *more* parameters as:

$$\text{Pseudo } R^2 = \frac{\text{random variance}_{\text{fewer}} - \text{random variance}_{\text{more}}}{\text{random variance}_{\text{fewer}}}$$

# Fixed Effects of Time-Invariant Predictors

- What about predicting level-1 effects with no random variance?
  - If the random linear time slope is n.s., can I test interactions with time?

**This should be ok to do...**

$\beta_{0i} = \gamma_{00} + \gamma_{01}Ed_i + U_{0i}$
$\beta_{1i} = \gamma_{10} + \gamma_{11}Ed_i + U_{1i}$
$\beta_{2i} = \gamma_{20} + \gamma_{21}Ed_i + U_{2i}$

**Is this still ok to do?**

$\beta_{0i} = \gamma_{00} + \gamma_{01}Ed_i + U_{0i}$
$\beta_{1i} = \gamma_{10} + \gamma_{11}Ed_i$
$\beta_{2i} = \gamma_{20} + \gamma_{21}Ed_i$

  - YES, surprisingly enough....
  - **In theory**, if a level-1 effect does not vary randomly over individuals, then it has "no" variance to predict (so cross-level interactions with that level-1 effect are not necessary)
  - However, because power to detect random effects is often lower than power to detect fixed effects, fixed effects of predictors can still be significant even if there is "no" (≈0) variance for them to predict
  - Just make sure you test for random effects BEFORE testing any cross-level interactions with that level-1 predictor!



# 3 Types of Effects: Fixed, Random, and Systematically (Non-Randomly) Varying

Let's say we have a significant fixed linear effect of time. What happens after we test a sex*time interaction?

| | Non-Significant Sex*Time effect? | Significant Sex*Time effect? |
|---|---|---|
| Random time slope initially **not** significant | Linear effect of time is **FIXED** | Linear effect of time is **systematically varying** |
| Random time initially sig, **not** sig. after sex*time | --- | Linear effect of time is **systematically varying** |
| Random time initially sig, **still** sig. after sex*time | Linear effect of time is **RANDOM** | Linear effect of time is **RANDOM** |

The effects of level-1 predictors (time-level) can be fixed, random, or systematically varying. The effects of level-2 predictors (person-level) can only be fixed or systematically varying (nothing to be random over...yet).

# Variance Accounted For By Level-2 Time-Invariant Predictors

- **Fixed effects of level 2 predictors *by themselves*:**
  - L2 (BP) main effects (e.g., sex) reduce L2 (BP) random intercept variance
  - L2 (BP) interactions (e.g., sex by ed) also reduce L2 (BP) random intercept variance

- **Fixed effects of *cross-level interactions* (level 1* level 2):**
  - If the interacting level-1 predictor is <u>random</u>, any cross-level interaction with it will reduce its corresponding level-2 BP random slope variance
    - e.g., if *time* is random, then sex**time*, ed**time*, and sex*ed**time* can each reduce the random linear time slope variance
  - If the interacting level-1 predictor <u>not random</u>, any cross-level interaction with it will reduce the level-1 WP residual variance instead
    - e.g., if $time^2$ is fixed, then sex*$time^2$, ed*$time^2$, and sex*ed*$time^2$ will reduce the L1 (WP) residual variance → Different quadratic slopes from sex and ed will allow better trajectories, reduce the variance around trajectories



# Variance Accounted for… For Real

- **Pseudo-$R^2$** is named that way for a reason… piles of variance can shift around, such that it can actually be negative
  - Sometimes a sign of model mis-specification
  - Hard to explain to readers when it happens!

- **One last simple alternative: Total $R^2$**
  - Generate model-predicted y's from fixed effects only (NOT including random effects) and correlate with observed y's
  - Then square correlation → total $R^2$
  - Total $R^2$ = total reduction in overall variance of y across levels
  - Can be "unfair" in models with large unexplained sources of variance

- MORAL OF THE STORY: Specify EXACTLY which kind of pseudo-$R^2$ you used—give the formula and the reference!!

# Time-Invariant Predictors in Longitudinal Models

- Topics:
  - Missing predictors in MLM
  - Effects of time-invariant predictors
  - Fixed, systematically varying, and random level-1 effects
  - **Model building strategies and assessing significance**



# Model-Building Strategies

- It may be helpful to examine predictor effects in separate models at first, including interactions with all growth terms to see the total pattern of effects for a single predictor
  - Question: Does age matter at all in predicting change over time?
  - e.g., random quadratic model + age, age*time, age*time$^2$
- Then predictor effects can be combined in layers in order to examine unique contributions (and interactions) of each
  - Question: Does age *still* matter after considering reasoning?
  - random quadratic + age, age*time, age*time$^2$,<br>+ reason, reason*time, reason*time$^2$
  - Potentially also + age*reason, age*reason*time, age*reason*time$^2$
- Sequence of predictors should be guided by theory and research questions—there may not be a single "best model"
  - One person's "control" is another person's "question", so may not end up in the same place given different orders of predictor inclusion

# Evaluating Statistical Significance of Multiple New Fixed Effects at Once

- Compare nested models with ML −2ΔLL test
- Useful for 'borderline' cases - example:
  - Ed*time$^2$ interaction at $p$ = .04, with nonsignificant ed*time and ed*Intercept (main effect of ed) terms?
  - Is it worth keeping a marginal higher-order interaction that requires two (possibly non-significant) lower-order terms?
  - ML −2ΔLL test on df=3: −2ΔLL must be > 7.82
  - **REML is WRONG for −2ΔLL tests for models with different fixed effects, regardless of nested or non-nested**
  - Because of this, it may be more convenient to switch to ML when focusing on modeling fixed effects of predictors
- Compare non-nested models with ML AIC & BIC instead



# Evaluating Statistical Significance of New Individual Fixed Effects

Fixed effects can be tested via **Wald** tests: the ratio of its estimate/SE forms a statistic we compare to a distribution

| | Denominator DF is assumed infinite | Denominator DF is estimated instead |
|---|---|---|
| Numerator DF = 1 | use ***z*** distribution (Mplus, STATA) | use ***t*** distribution (SAS, SPSS) |
| Numerator DF > 1 | use $\chi^2$ distribution (Mplus, STATA) | use ***F*** distribution (SAS, SPSS) |
| Denominator DF (DDFM) options | not applicable, so DDF is not given | SAS: BW and KR SAS and SPSS: Satterthwaite |

# Denominator DF (DDF) Methods

- **Between-Within** (DDFM=BW in SAS, not in SPSS):
  - Total DDF (T) comes from total number of observations, separated into level-2 for *N* persons and level-1 for *n* occasions
    - **Level-2 DDF** = *N* – #level-2 fixed effects
    - **Level-1 DDF** = Total DDF – Level-2 DDF – #level-1 fixed effects
    - Level-1 effects with random slopes still get level-1 DDF
- **Satterthwaite** (DDFM=Satterthwaite in SAS, default in SPSS):
  - More complicated, but analogous to two-group *t*-test given unequal residual variances and unequal group sizes
  - Incorporates contribution of variance components at each level
    - Level-2 DDF will resemble Level-2 DDF from BW
    - Level-1 DDF will resemble Level-1 DDF from BW if the level-1 effect is not random, but will resemble level-2 DDF if it is random



# Denominator DF (DDF) Methods

- **Kenward-Roger** (DDFM=KR in SAS, not in SPSS):
  - Adjusts the sampling covariance matrix of the fixed effects and variance components to reflect the uncertainty introduced by using large-sample techniques of ML/REML in small *N* samples
  - This creates different (larger) SEs for the fixed effects
  - Then uses Satterthwaite DDF, new SEs, and *t* to get *p*-values
- In an unstructured variance model, all effects use level-2 DDF
- Differences in inference not likely to matter often in practice
  - e.g., critical *t*-value at DDF=20 is 2.086, at infinite DDF is 1.960
- When in doubt, use KR (is overkill at worst, becomes Satterthwaite)
  - I used Satterthwaite in the book to maintain comparability across programs

# Wrapping Up…

- MLM uses ONLY rows of data that are COMPLETE: both predictors AND outcomes must be there!
  - Using whatever data you do have for each person will likely lead to better inferences and more statistical power than using only complete persons (listwise deletion)
- Time-invariant predictors modify the level-1 created growth curve → predict individual intercepts and slopes
  - They account for random effect variances (the predictors are the reasons WHY people need their own intercepts and slopes)
  - If a level-1 effect is not random, it can still be moderated by a cross-level interaction with a time-invariant predictor…
    - … but then it will predict L1 residual variance instead

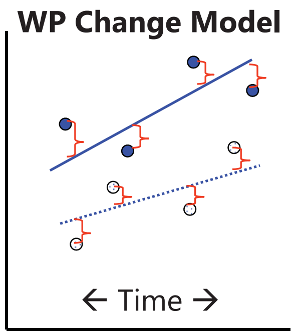


# Time-Varying Predictors in Longitudinal Models

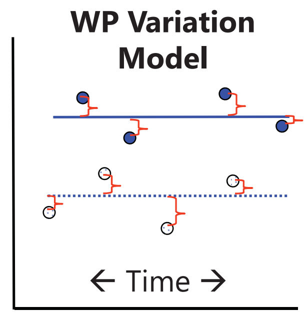
- Topics:
  - **Time-varying predictors that fluctuate over time**
  - Person-Mean-Centering (PMC)
  - Grand-Mean-Centering (GMC)
  - Model extensions under Person-MC vs. Grand-MC
  - Time-varying predictors that change over time

# The Joy of Time-Varying Predictors

- TV predictors predict leftover **WP (residual) variation**:




If model for time works, then residuals should look like this →




- Modeling time-varying predictors is complicated because they represent an **aggregated effect**:
  - Effect of the *between-person* variation in the predictor $x_{ti}$ on Y
  - Effect of the *within-person* variation in the predictor $x_{ti}$ on Y
  - Here we are assuming the predictor $x_{ti}$ only **fluctuates** over time…
    - *We will need a different model if $x_{ti}$ changes systematically over time…*



# The Joy of Time-Varying Predictors

- Time-varying (TV) predictors usually carry 2 kinds of effects because they are really 2 predictor variables, not 1

- Example: Stress measured daily
  - Some days are worse than others:
    - **WP variation in stress** *(represented as deviation from own mean)*
  - Some people just have more stress than others all the time:
    - **BP variation in stress** *(represented as person mean predictor over time)*

- Can quantify each source of variation with an ICC
  - ICC = (BP variance) / (BP variance + WP variance)
  - ICC > 0? TV predictor has BP variation (so it *could* have a BP effect)
  - ICC < 1? TV predictor has WP variation (so it *could* have a WP effect)

# Between-Person vs. Within-Person Effects

- Between-person and within-person effects in SAME direction
  - Stress → Health?
    - **BP: People with more chronic stress than other people may have worse general health than people with less chronic stress**
    - **WP: People may feel worse than usual when they are currently under more stress than usual (regardless of what "usual" is)**
- Between-person and within-person effects in OPPOSITE directions
  - Exercise → Blood pressure?
    - **BP: People who exercise more often generally have lower blood pressure than people who are more sedentary**
    - **WP: During exercise, blood pressure is higher than during rest**
- Variables have different **meanings** at different levels!
- Variables have different **scales** at different levels



# 3 Kinds of Effects for TV Predictors

- **Is the Between-Person (BP) effect significant?**
  - Are people with higher predictor values than other people (*on average over time*) also higher on Y than other people (*on average over time*), such that the person mean of the TV predictor accounts for level-2 random intercept variance ($\tau^2_{U_0}$)?
- **Is the Within-Person (WP) effect significant?**
  - If you have higher predictor values than usual (*at this occasion*), do you also have higher outcomes values than usual (*at this occasion*), such that the within-person deviation of the TV predictor accounts for level-1 residual variance ($\sigma^2_e$)?
- **Are the BP and WP effects different sizes: Is there a contextual effect?**
  - After controlling for the absolute value of TV predictor at each occasion, is there still an incremental contribution from having a higher person mean of the TV predictor (i.e., does one's general tendency predict $\tau^2_{U_0}$ above and beyond)?
  - If there is no contextual effect, then the BP and WP effects of the TV predictor show *convergence*, such that their effects are of equivalent magnitude

# Modeling TV Predictors (labeled as $x_{ti}$)

- **Level-2 effect of $x_{ti}$:**
  - The level-2 effect of $x_{ti}$ is usually represented by the person's mean of time-varying $x_{ti}$ across time (labeled as **PMx$_i$** or $\bar{\mathbf{X}}_\mathbf{i}$)
  - PMx$_i$ should be centered at a CONSTANT (grand mean or other) so that 0 is meaningful, just like any other time-invariant predictor

- **Level-1 effect of $x_{ti}$ can be included two different ways:**
  - "**Group-mean-centering**" → "**person-mean-centering**" in longitudinal, in which level-1 predictors are centered using a level-2 VARIABLE
  - "**Grand-mean-centering**" → level-1 predictors are centered using a CONSTANT (not necessarily the grand mean; it's just called that)
  - Note that these 2 choices do NOT apply to the level-2 effect of $x_{ti}$!
    - But the interpretation of the level-2 effect of $x_{ti}$ WILL DIFFER based on which centering method you choose for the level-1 effect of $x_{ti}$!



# Time-Varying Predictors in Longitudinal Models

- Topics:
  - Time-varying predictors that fluctuate over time
  - **Person-Mean-Centering (PMC)**
  - Grand-Mean-Centering (GMC)
  - Model extensions under Person-MC vs. Grand-MC
  - Time-varying predictors that change over time

# Person-Mean-Centering (P-MC)

- In P-MC, we decompose the TV predictor $x_{ti}$ into 2 variables that directly represent its BP (level-2) and WP (level-1) sources of variation, and include those variables as the predictors instead:
- **Level-2, PM predictor = person mean of $x_{ti}$**
  - $\mathbf{PMx_i} = \mathbf{\bar{X}_i} - \boldsymbol{C}$
  - $PMx_i$ is centered at a constant $C$, chosen so 0 is meaningful
  - $PMx_i$ is positive? Above sample mean → "more than other people"
  - $PMx_i$ is negative? Below sample mean → "less than other people"
- **Level-1, WP predictor = deviation from person mean of $x_{ti}$**
  - $\mathbf{WPx_{ti}} = \mathbf{x_{ti}} - \mathbf{\bar{X}_i}$ *(note: uncentered person mean $\bar{X}_i$ is used to center $x_{ti}$)*
  - $WPx_{ti}$ is NOT centered at a constant; is centered at a VARIABLE
  - $WPx_{ti}$ is positive? Above your own mean → "more than usual"
  - $WPx_{ti}$ is negative? Below your own mean → "less than usual"



# Within-Person Fluctuation Model with **Person-Mean-Centered Level-1 $x_{ti}$**

## → WP and BP Effects directly through **separate** parameters

**$x_{ti}$ is person-mean-centered into $WPx_{ti}$, with $PMx_i$ at L2:**

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(WPx_{ti}) + e_{ti}$

**$WPx_{ti} = x_{ti} - \bar{X}_i$ → it has only Level-1 WP variation**

Level 2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(PMx_i) + U_{0i}$

$\beta_{1i} = \gamma_{10}$

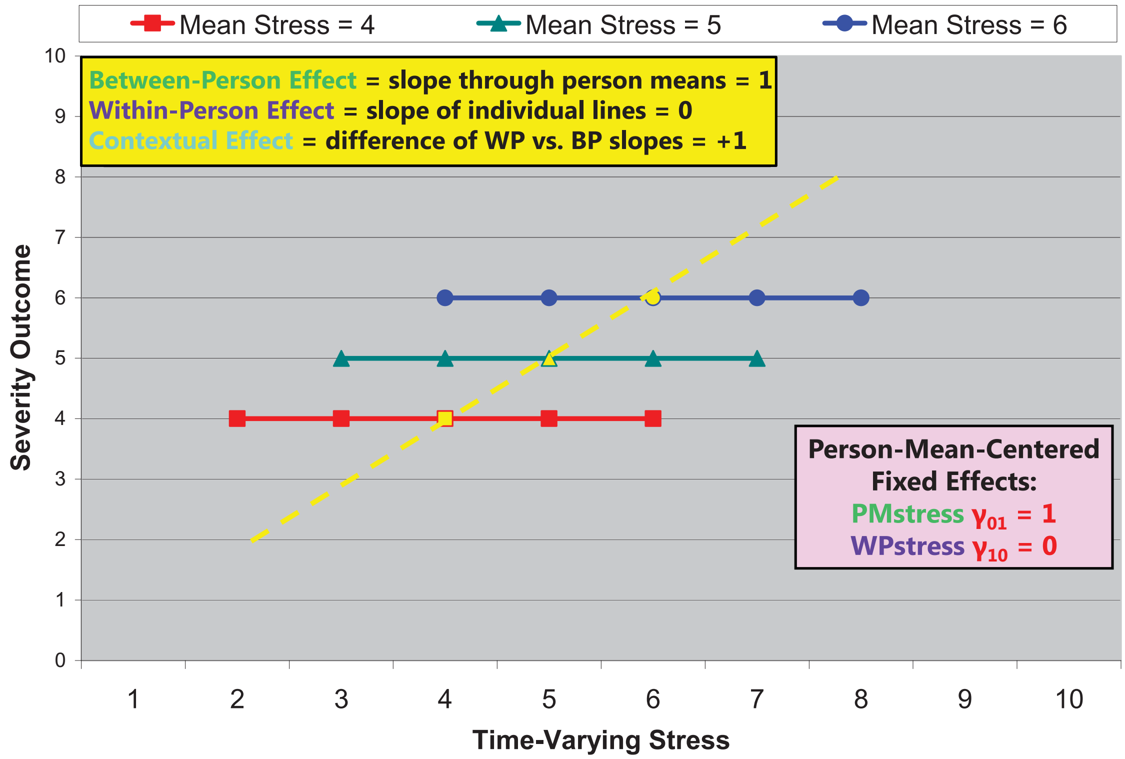
**$PMx_i = \bar{X}_i - C$ → it has only Level-2 BP variation**

**$\gamma_{10}$ = WP main effect of having more $x_{ti}$ than usual**

**$\gamma_{01}$ = BP main effect of having more $\bar{X}_i$ than other people**

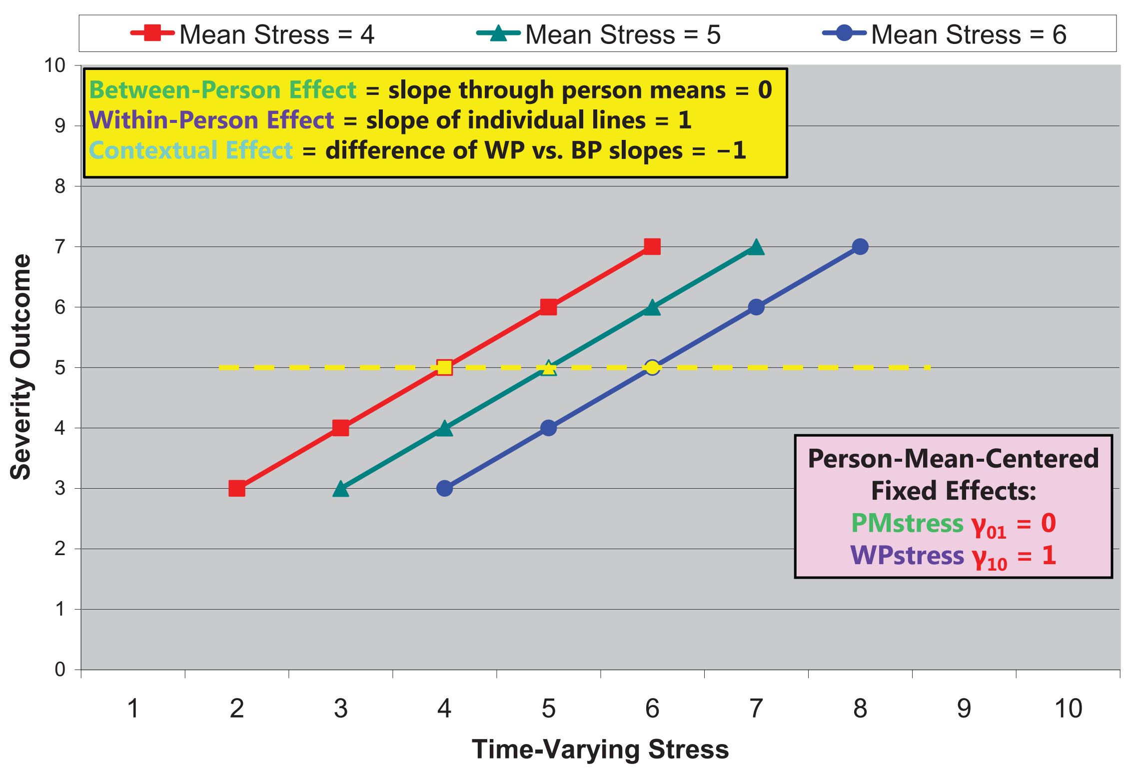
**Because $WPx_{ti}$ and $PMx_i$ are uncorrelated, each gets the <u>total</u> effect for its level (WP=L1, BP=L2)**

# ALL Between-Person Effect, NO Within-Person Effect

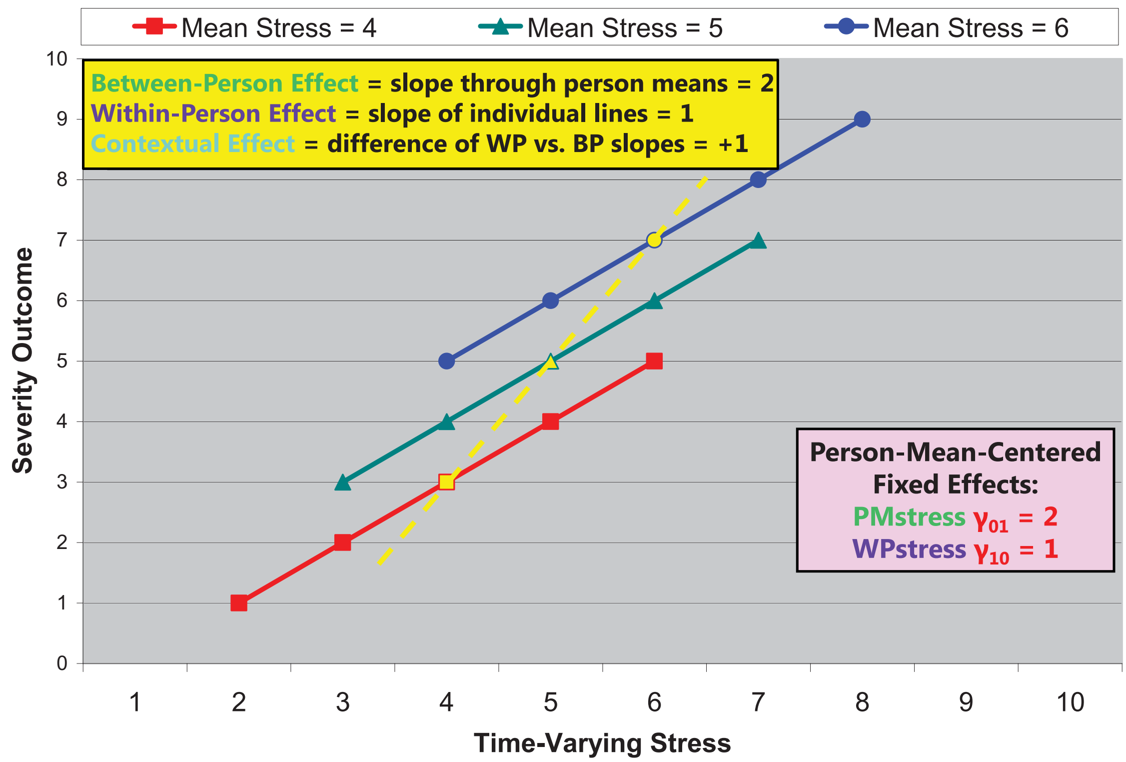





# NO Between-Person Effect, ALL Within-Person Effect



## Between-Person Effect > Within-Person Effect






## Within-Person Fluctuation Model with **Person-Mean-Centered Level-1 $x_{ti}$**

→ WP and BP Effects directly through **separate** parameters

**$x_{ti}$ is person-mean-centered into $WPx_{ti}$, with $PMx_i$ at L2:**

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(WPx_{ti}) + e_{ti}$

**$WPx_{ti} = x_{ti} - \bar{X}_i$ → it has only Level-1 WP variation**

Level 2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(PMx_i) + U_{0i}$

$\beta_{1i} = \gamma_{10} + \gamma_{11}(PMx_i) + U_{1i}$

**$PMx_i = \bar{X}_i - C$ → it has only Level-2 BP variation**

**$U_{1i}$ is a random slope for the WP effect of $x_{ti}$**

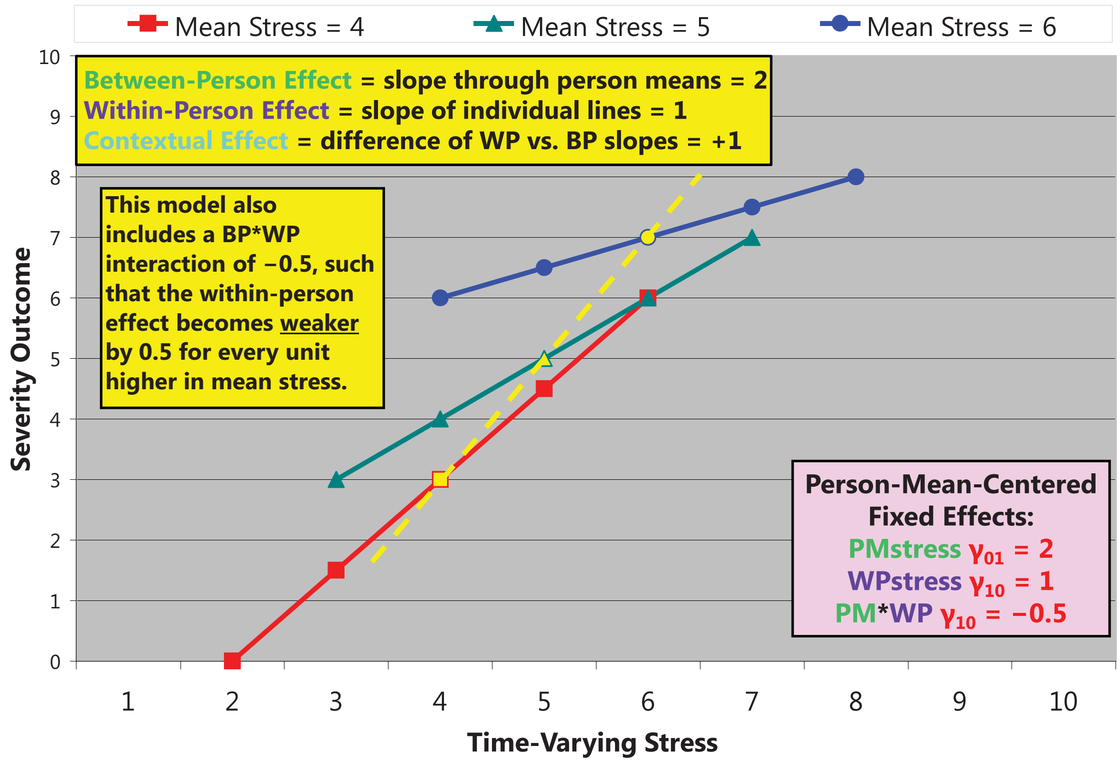
**$\gamma_{10}$ = WP simple main effect of having more $x_{ti}$ than usual *for* $PMx_i = 0$**

**$\gamma_{01}$ = BP simple main effect of having more $\bar{X}_i$ than other people for people at their own mean ($WPx_{ti} = x_{ti} - \bar{X}_i \rightarrow 0$)**

**$\gamma_{11}$ = BP*WP interaction: how the effect of having more $x_{ti}$ than usual differs by how much $\bar{X}_i$ you have**

Note: this model should also test $\gamma_{02}$ for $PMx_i * PMxi$ (stay tuned)

# Between-Person x Within-Person Interaction






# Time-Varying Predictors in Longitudinal Models

- Topics:
  - Time-varying predictors that fluctuate over time
  - Person-Mean-Centering (PMC)
  - **Grand-Mean-Centering (GMC)**
  - Model extensions under Person-MC vs. Grand-MC
  - Time-varying predictors that change over time

# 3 Kinds of Effects for TV Predictors

- **What Person-Mean-Centering tells us directly:**

- **Is the Between-Person (BP) effect significant?**
  - Are people with higher predictor values than other people (*on average over time*) also higher on Y than other people (*on average over time*), such that the person mean of the TV predictor accounts for level-2 random intercept variance ($\tau^2_{U_0}$)?
  - This would be indicated by a significant fixed effect of $\mathbf{PMx_i}$
  - Note: this is NOT controlling for the absolute value of $x_{ti}$ at each occasion

- **Is the Within-Person (WP) effect significant?**
  - If you have higher predictor values than usual (*at this occasion*), do you also have higher outcomes values than usual (*at this occasion*), such that the within-person deviation of the TV predictor accounts for level-1 residual variance ($\sigma^2_e$)?
  - This would be indicated by a significant fixed effect of $\mathbf{WPx_{ti}}$
  - Note: this is represented by the relative value of $x_{ti}$, NOT the absolute value of $x_{ti}$



# 3 Kinds of Effects for TV Predictors

- **What Person-Mean-Centering DOES NOT tell us directly:**

- **Are the BP and WP effects different sizes: Is there a contextual effect?**
  - After controlling for the absolute value of the TV predictor at each occasion, is there still an incremental contribution from having a higher person mean of the TV predictor (i.e., does one's general tendency predict $\tau^2_{U_0}$ above and beyond just the time-specific value of the predictor)?
  - If there is no contextual effect, then the BP and WP effects of the TV predictor show ***convergence***, such that their effects are of equivalent magnitude

- **To answer this question about the contextual effect for the incremental contribution of the person mean, we have two options:**
  - Ask for the contextual effect via an ESTIMATE statement in SAS (or TEST in SPSS, or NEW in Mplus, or LINCOM in STATA): `WPx_ti -1 PMx_i 1`
  - Use "**grand-mean-centering**" for time-varying $x_{ti}$ instead: $\mathbf{TVx_{ti}} = \mathbf{x_{ti}} - C$ → **centered at a CONSTANT, NOT A LEVEL-2 VARIABLE**
    - Which constant only matters for what the reference point is; it could be the grand mean or other

# Remember Regular Old Regression?

- In this model: $Y_i = \beta_0 + \beta_1(X_{1i}) + \beta_2(X_{2i}) + e_i$
  - If $X_{1i}$ and $X_{2i}$ **ARE NOT** correlated:
    - $\beta_1$ is **ALL the relationship** between $X_{1i}$ and $Y_i$
    - $\beta_2$ is **ALL the relationship** between $X_{2i}$ and $Y_i$
  - If $X_{1i}$ and $X_{2i}$ **ARE** correlated:
    - $\beta_1$ is **different than** the full relationship between $X_{1i}$ and $Y_i$
      - "Unique" effect of $X_{1i}$ *controlling for* $X_{2i}$ or *holding* $X_{2i}$ *constant*
    - $\beta_2$ is **different than** the full relationship between $X_{2i}$ and $Y_i$
      - "Unique" effect of $X_{2i}$ *controlling for* $X_{1i}$ or *holding* $X_{1i}$ *constant*
  - Hang onto that idea…



# Person-MC vs. Grand-MC for Time-Varying Predictors

| Level 2 | | Original | Person-MC Level 1 | Grand-MC Level 1 |
|---|---|---|---|---|
| $\bar{X}_i$ | $PMx_i = \bar{X}_i - 5$ | xti | $WPx_{ti} = x_{ti} - \bar{X}_i$ | $TVx_{ti} = x_{ti} - 5$ |
| 3 | −2 | 2 | −1 | −3 |
| 3 | −2 | 4 | 1 | −1 |
| 7 | 2 | 6 | −1 | 1 |
| 7 | 2 | 8 | 1 | 3 |

Same $PMx_i$ goes into the model using either way of centering the level-1 variable $x_{ti}$

Using **Person-MC**, $WPx_{ti}$ has NO level-2 BP variation, so it is not correlated with $PMx_i$

Using **Grand-MC**, $TVx_{ti}$ STILL has level-2 BP variation, so it is STILL CORRELATED with $PMx_i$

**So the effects of $PMx_i$ and $TVx_{ti}$ when included together under Grand-MC will be different than their effects would be if they were by themselves…**

# Within-Person Fluctuation Model with $\mathbf{x_{ti}}$ represented at Level 1 Only:

## → WP and BP Effects are **Smushed Together**

**$x_{ti}$ is grand-mean-centered into $TVx_{ti}$, WITHOUT $PMx_i$ at L2:**

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(\mathbf{TVx_{ti}}) + e_{ti}$

**$TVx_{ti} = x_{ti} - C$ → it still has both Level-2 BP and Level-1 WP variation**

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

$\beta_{1i} = \gamma_{10}$

**$\gamma_{10}$ = *smushed* WP and BP effects**

**Because $TVx_{ti}$ still contains its original 2 different kinds of variation (BP and WP), its 1 fixed effect has to do the work of 2 predictors!**

**A *smushed* effect is also referred to as the *convergence*, *conflated*, or *composite* effect**



# Convergence (Smushed) Effect of a Time-Varying Predictor

$$\text{Convergence Effect: } \gamma_{conv} \approx \frac{\dfrac{\gamma_{BP}}{SE^2_{BP}} + \dfrac{\gamma_{WP}}{SE^2_{WP}}}{\dfrac{1}{SE^2_{BP}} + \dfrac{1}{SE^2_{WP}}}$$

Adapted from Raudenbush & Bryk (2002, p. 138)

- **The convergence effect will often be closer to the within-person effect** (due to larger level-1 sample size and thus smaller SE)
- **It is the rule, not the exception, that between and within effects differ** (Snijders & Bosker, 1999, p. 52-56, and personal experience!)
- However—when grand-mean-centering a time-varying predictor, **convergence is testable** by including a **contextual effect (carried by the person mean)** for how the **BP effect** differs from the **WP effect**…

# Within-Person Fluctuation Model with **Grand-Mean-Centered Level-1 $x_{ti}$**

→ Model tests difference of WP vs. BP effects (It's been fixed!)

**$x_{ti}$ is grand-mean-centered into $TVx_{ti}$, WITH $PMx_i$ at L2:**

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(TVx_{ti}) + e_{ti}$

$TVx_{ti} = x_{ti} - C$ → **it still has both Level-2 BP and Level-1 WP variation**

Level 2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(PMx_i) + U_{0i}$

$\beta_{1i} = \gamma_{10}$

$PMx_i = \bar{X}_i - C$ → **it has only Level-2 BP variation**

**$\gamma_{10}$ becomes the WP effect → *unique* level-1 effect after controlling for $PMx_i$**

**$\gamma_{01}$ *becomes* the contextual effect that indicates how the BP effect differs from the WP effect**
**→ *unique* level-2 effect after controlling for $TVx_{ti}$**
**→ does usual level matter beyond current level?**



# Person-MC and Grand-MC Models are Equivalent Given a Fixed Level-1 Main Effect Only

**Person-MC:** $WPx_{ti} = x_{ti} - PMx_i$

**Level-1:** $y_{ti} = \beta_{0i} + \beta_{1i}(x_{ti} - PMx_i) + e_{ti}$

**Level-2:** $\beta_{0i} = \gamma_{00} + \gamma_{01}(PMx_i) + U_{0i}$

$\beta_{1i} = \gamma_{10}$

→ $y_{ti} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{10}(x_{ti} - PMx_i) + U_{0i} + e_{ti}$

→ $y_{ti} = \gamma_{00} + (\gamma_{01} - \gamma_{10})(PMx_i) + \gamma_{10}(x_{ti}) + U_{0i} + e_{ti}$

**Composite Model:**
**← In terms of P-MC**
**← In terms of G-MC**

---

**Grand-MC:** $TVx_{ti} = x_{ti}$

**Level-1:** $y_{ti} = \beta_{0i} + \beta_{1i}(x_{ti}) + e_{ti}$

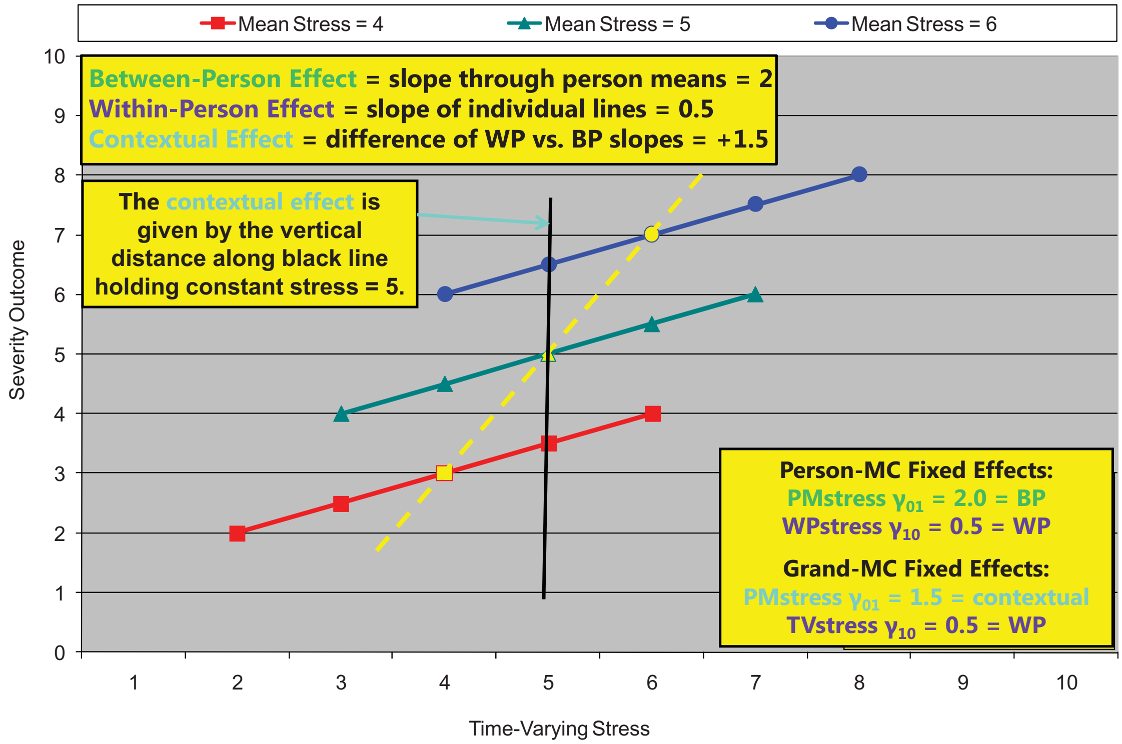
**Level-2:** $\beta_{0i} = \gamma_{00} + \gamma_{01}(PMx_i) + U_{0i}$

$\beta_{1i} = \gamma_{10}$

→ $y_{ti} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{10}(x_{ti}) + U_{0i} + e_{ti}$

| Effect | P-MC | G-MC |
|---|---|---|
| Intercept | $\gamma_{00}$ | $\gamma_{00}$ |
| WP Effect | $\gamma_{10}$ | $\gamma_{10}$ |
| Contextual | $\gamma_{01} - \gamma_{10}$ | $\gamma_{01}$ |
| BP Effect | $\gamma_{01}$ | $\gamma_{01} + \gamma_{10}$ |

# P-MC vs. G-MC: Interpretation Example






# Summary: 3 Effects for TV Predictors

- **Is the Between-Person (BP) effect significant?**
  - Are people with higher predictor values than other people (*on average over time*) also higher on Y than other people (*on average over time*), such that the person mean of the TV predictor accounts for level-2 random intercept variance ($\tau^2_{U_0}$)?
  - Given directly by level-2 effect of $PMx_i$ if using Person-MC for the level-1 predictor (or can be requested via ESTIMATE if using Grand-MC for the level-1 predictor)
- **Is the Within-Person (WP) effect significant?**
  - If you have higher predictor values than usual (*at this occasion*), do you also have higher outcomes values than usual (*at this occasion*), such that the within-person deviation of the TV predictor accounts for level-1 residual variance ($\sigma^2_e$)?
  - Given directly by the level-1 effect of $WPx_{ti}$ if using Person-MC —*OR* — given directly by the level-1 effect of $TVx_{ti}$ if using Grand-MC and including $PMx_i$ at level 2 (without $PMx_i$, the level-1 effect of $TVx_{ti}$ if using Grand-MC is the smushed effect)
- **Are the BP and WP Effects different sizes: Is there a contextual effect?**
  - After controlling for the absolute value of TV predictor value at each occasion, is there still an incremental contribution from having a higher person mean of the TV predictor (i.e., does one's general tendency predict $\tau^2_{U_0}$ above and beyond)?
  - Given directly by level-2 effect of $PMx_i$ if using Grand-MC for the level-1 predictor (or can be requested via ESTIMATE if using Person-MC for the level-1 predictor)

# Variance Accounted For By Level-1 Predictors

- **Fixed effects of level 1 predictors *by themselves*:**
  - Level-1 (WP) main effects reduce Level-1 (WP) residual variance
  - Level-1 (WP) interactions also reduce Level-1 (WP) residual variance

- **What happens at level 2 depends on what kind of variance the level-1 predictor has:**
  - If the level-1 predictor ALSO has level-2 variance (e.g., Grand-MC predictors), then its level-2 variance will also likely reduce level-2 random intercept variance
  - If the level-1 predictor DOES NOT have level-2 variance (e.g., Person-MC predictors), then its reduction in the level-1 residual variance will cause an INCREASE in level-2 random intercept variance
    - Same thing happens with Grand-MC level-1 predictors, but you don't generally see it
  - It's just an artifact that the estimate of true random intercept variance is:
    True $\tau^2_{U_0}$ = observed $\tau^2_{U_0} - \frac{\sigma^2_e}{n}$ → so if only $\sigma^2_e$ decreases, $\tau^2_{U_0}$ increases



# Time-Varying Predictors in Longitudinal Models

- Topics:
  - Time-varying predictors that fluctuate over time
  - Person-Mean-Centering (PMC)
  - Grand-Mean-Centering (GMC)
  - **Model extensions under Person-MC vs. Grand-MC**
  - Time-varying predictors that change over time

# The Joy of Interactions Involving Time-Varying Predictors

- Must consider interactions with both its BP and WP parts:
- Example: Does time-varying stress ($x_{ti}$) interact with sex ($Sex_i$)?

- Person-Mean-Centering:
  - $WPx_{ti} * Sex_i$ → Does the WP stress effect differ between men and women?
  - $PMx_i * Sex_i$ → Does the BP stress effect differ between men and women?
    - Not controlling for current levels of stress
    - If forgotten, then $Sex_i$ moderates the stress effect only at level 1 (WP, not BP)

- Grand-Mean-Centering:
  - $TVx_{ti} * Sex_i$ → Does the WP stress effect differ between men and women?
  - $PMx_i * Sex_i$ → Does the *contextual* stress effect differ b/t men and women?
    - Incremental BP stress effect *after controlling for current levels of stress*
    - If forgotten, then although the level-1 main effect of stress has been un-smushed via the main effect of $PMx_i$, the interaction of $TVx_{ti} * Sex_i$ would still be smushed



# Interactions with Time-Varying Predictors: Example: TV Stress ($x_{ti}$) by Gender ($Sex_i$)

**Person-MC:** $WPx_{ti} = x_{ti} - PMx_i$

**Level-1:** $y_{ti} = \beta_{0i} + \beta_{1i}(x_{ti} - PMx_i) + e_{ti}$

**Level-2:** $\beta_{0i} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{02}(Sex_i) + \gamma_{03}(Sex_i)(PMx_i) + U_{0i}$

$\beta_{1i} = \gamma_{10} + \gamma_{11}(Sex_i)$

**Composite:** $y_{ti} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{10}(x_{ti} - PMx_i) + U_{0i} + e_{ti}$
$+ \gamma_{02}(Sex_i) + \gamma_{03}(Sex_i)(PMx_i) + \gamma_{11}(Sex_i)(x_{ti} - PMx_i)$

---

**Grand-MC:** $TVx_{ti} = x_{ti}$

**Level-1:** $y_{ti} = \beta_{0i} + \beta_{1i}(x_{ti}) + e_{ti}$

**Level-2:** $\beta_{0i} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{02}(Sex_i) + \gamma_{03}(Sex_i)(PMx_i) + U_{0i}$

$\beta_{1i} = \gamma_{10} + \gamma_{11}(Sex_i)$

**Composite:** $y_{ti} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{10}(x_{ti}) + U_{0i} + e_{ti}$
$+ \gamma_{02}(Sex_i) + \gamma_{03}(Sex_i)(PMx_i) + \gamma_{11}(Sex_i)(x_{ti})$

# Interactions Involving Time-Varying Predictors Belong at Both Levels of the Model

**On the left below → Person-MC:** $WPx_{ti} = x_{ti} - PMx_i$

$y_{ti} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{10}(x_{ti} - PMx_i) + U_{0i} + e_{ti}$
$+ \gamma_{02}(Sex_i) + \gamma_{03}(Sex_i)(PMx_i) + \gamma_{11}(Sex_i)(x_{ti} - PMx_i)$

← **Composite model written as Person-MC**

$y_{ti} = \gamma_{00} + (\gamma_{01} - \gamma_{10})(PMx_i) + \gamma_{10}(x_{ti}) + U_{0i} + e_{ti}$
$+ \gamma_{02}(Sex_i) + (\gamma_{03} - \gamma_{11})(Sex_i)(PMx_i) + \gamma_{11}(Sex_i)(x_{ti})$

← **Composite model written as Grand-MC**

**On the right below → Grand-MC:** $TVx_{ti} = x_{ti}$

$y_{ti} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{10}(x_{ti}) + U_{0i} + e_{ti}$
$+ \gamma_{02}(Sex_i) + \gamma_{03}(Sex_i)(PMx_i) + \gamma_{11}(Sex_i)(x_{ti})$

After adding an interaction for $Sex_i$ with stress at both levels, then the Person-MC and Grand-MC models are equivalent

**Intercept:** $\gamma_{00} = \gamma_{00}$ **BP Effect:** $\gamma_{01} = \gamma_{01} + \gamma_{10}$ **Contextual:** $\gamma_{01} = \gamma_{01} - \gamma_{10}$

**WP Effect:** $\gamma_{10} = \gamma_{10}$ **BP*Sex Effect:** $\gamma_{03} = \gamma_{03} + \gamma_{11}$ **Contextual*Sex:** $\gamma_{03} = \gamma_{03} - \gamma_{11}$

**Sex Effect:** $\gamma_{20} = \gamma_{20}$ **BP*WP or Contextual*WP is the same:** $\gamma_{11} = \gamma_{11}$



# Intra-variable Interactions

- Still must consider interactions with both its BP and WP parts!
- Example: Interaction of TV stress ($x_{ti}$) with person mean stress ($PMx_i$)
- Person-Mean-Centering:
  - $WPx_{ti} * PMx_i$ → Does the WP stress effect differ by overall stress level?
  - $PMx_i * PMx_i$ → Does the BP stress effect differ by overall stress level?
    - Not controlling for current levels of stress
    - If forgotten, then $PMx_i$ moderates the stress effect only at level 1 (WP, not BP)
- Grand-Mean-Centering:
  - $TVx_{ti} * PMx_i$ → Does the WP stress effect differ by overall stress level?
  - $PMx_i * PMx_i$ → Does the *contextual* stress effect differ by overall stress?
    - Incremental BP stress effect *after controlling for current levels of stress*
    - If forgotten, then although the level-1 main effect of stress has been un-smushed via the main effect of $PMx_i$, the interaction of $TVx_{ti} * PMx_i$ would still be smushed

# Intra-variable Interactions: Example: TV Stress ($x_{ti}$) by Person Mean Stress ($PMx_i$)

**Person-MC:** $WPx_{ti} = x_{ti} - PMx_i$

**Level-1:** $y_{ti} = \beta_{0i} + \beta_{1i}(x_{ti} - PMx_i) + e_{ti}$

**Level-2:** $\beta_{0i} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{02}(PMx_i)(PMx_i) + U_{0i}$

$\beta_{1i} = \gamma_{10} + \gamma_{11}(PMx_i)$

**Composite:** $y_{ti} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{10}(x_{ti} - PMx_i) + U_{0i} + e_{ti} + \gamma_{02}(PMx_i)(PMx_i) + \gamma_{11}(PMx_i)(x_{ti} - PMx_i)$

---

**Grand-MC:** $TVx_{ti} = x_{ti}$

**Level-1:** $y_{ti} = \beta_{0i} + \beta_{1i}(x_{ti}) + e_{ti}$

**Level-2:** $\beta_{0i} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{02}(PMx_i)(PMx_i) + U_{0i}$

$\beta_{1i} = \gamma_{10} + \gamma_{11}(PMx_i)$

**Composite:** $y_{ti} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{10}(x_{ti}) + U_{0i} + e_{ti} + \gamma_{02}(PMx_i)(PMx_i) + \gamma_{11}(PMx_i)(x_{ti})$



# Intra-variable Interactions: Example: TV Stress ($x_{ti}$) by Person Mean Stress ($PMx_i$)

**On the left below → Person-MC:** $WPx_{ti} = x_{ti} - PMx_i$

$y_{ti} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{10}(x_{ti} - PMx_i) + U_{0i} + e_{ti} + \gamma_{02}(PMx_i)(PMx_i) + \gamma_{11}(PMx_i)(x_{ti} - PMx_i)$

**← Written as Person-MC**

$y_{ti} = \gamma_{00} + (\gamma_{01} - \gamma_{10})(PMx_i) + \gamma_{10}(x_{ti}) + U_{0i} + e_{ti} + (\gamma_{02} - \gamma_{11})(PMx_i)(PMx_i) + \gamma_{11}(PMx_i)(x_{ti})$

**← Written as Grand-MC**

---

**On the right below → Grand-MC:** $TVx_{ti} = x_{ti}$

$y_{ti} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{10}(x_{ti}) + U_{0i} + e_{ti} + \gamma_{02}(PMx_i)(PMx_i) + \gamma_{11}(PMx_i)(x_{ti})$

After adding an interaction for $PMx_i$ with stress at both levels, then the Person-MC and Grand-MC models are equivalent

---

**Intercept:** $\gamma_{00} = \gamma_{00}$ **BP Effect:** $\gamma_{01} = \gamma_{01} + \gamma_{10}$ **Contextual:** $\gamma_{01} = \gamma_{01} - \gamma_{10}$

**WP Effect:** $\gamma_{10} = \gamma_{10}$ **BP² Effect:** $\gamma_{02} = \gamma_{02} + \gamma_{11}$ **Contextual²:** $\gamma_{02} = \gamma_{02} - \gamma_{11}$

**BP*WP or Contextual*WP is the same:** $\gamma_{11} = \gamma_{11}$

# When Person-MC ≠ Grand-MC: Random Effects of TV Predictors

**Person-MC:** $WPx_{ti} = x_{ti} - PMx_i$

Level-1: $y_{ti} = \beta_{0i} + \beta_{1i}(x_{ti} - PMx_i) + e_{ti}$

Level-2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(PMx_i) + U_{0i}$

$\beta_{1i} = \gamma_{10} + U_{1i}$

$\rightarrow y_{ti} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{10}(x_{ti} - PMx_i) + U_{0i} + U_{1i}(x_{ti} - PMx_i) + e_{ti}$

Variance due to $PMx_i$ is removed from the random slope in Person-MC.

**Grand-MC:** $TVx_{ti} = x_{ti}$

Level-1: $y_{ti} = \beta_{0i} + \beta_{1i}(x_{ti}) + e_{ti}$

Level-2: $\beta_{0i} = \gamma_{00} + \gamma_{01}(PMx_i) + U_{0i}$

$\beta_{1i} = \gamma_{10} + U_{1i}$

$\rightarrow y_{ti} = \gamma_{00} + \gamma_{01}(PMx_i) + \gamma_{10}(x_{ti}) + U_{0i} + U_{1i}(x_{ti}) + e_{ti}$

Variance due to $PMx_i$ is still part of the random slope in Grand-MC. So these models cannot be made equivalent.



# Random Effects of TV Predictors

- **Random intercepts** mean different things under each model:
  - **Person-MC** → Individual differences at **$WPx_{ti}$ =0** (that everyone has)
  - **Grand-MC** → Individual differences at **$TVx_{ti}$=0** (that not everyone has)
- **Differential shrinkage of the random intercepts** results from differential reliability of the intercept data across models:
  - Person-MC → Won't affect shrinkage of slopes unless highly correlated
  - Grand-MC → Will affect shrinkage of slopes due to forced extrapolation
- As a result, the **random slope variance may be too small** when using Grand-MC rather than Person-MC
  - Problem worsens with greater ICC of TV Predictor (more extrapolation)
  - Anecdotal example using clustered data was presented in Raudenbush & Bryk (2002; chapter 6)

# Modeling Time-Varying Categorical Predictors

- Person-MC and Grand-MC really only apply to *continuous* TV predictors, but the need to consider BP and WP effects applies to *categorical* TV predictors too
- Binary level-1 predictors do not lend themselves to Person-MC
  - e.g., $x_{ti}$ = 0 or 1 per occasion, person mean = .50 across occasions → impossible values
  - If $x_{ti} = 0$, then $WPx_{ti} = 0 - .50 = -0.50$; If $x_{ti} = 1$, then $WPx_{ti} = 1 - .50 = 0.50$
  - Better: Leave $x_{ti}$ uncentered and include person mean as level-2 predictor (results ~ Grand-MC)
- For >2 categories, person means of multiple dummy codes starts to break down, but we can think about types of people, and code BP effects accordingly
- Example: Dementia present/not at each time point?
  - **BP effects** → Ever diagnosed with dementia (no, yes)?
    - People who will eventually be diagnosed may differ prior to diagnosis (a BP effect)
  - **TV effect** → Diagnosed with dementia at each time point (no, yes)?
    - Acute differences of before/after diagnosis logically can only exist in the "ever" people
- Other examples: Mentor status, father absence, type of shift work (AM/PM)



# Wrapping Up: Person-MC vs. Grand-MC

- Time-varying predictors carry at least two potential effects:
  - Some people are higher/lower than other people → BP, level-2 effect
  - Some occasions are higher/lower than usual → WP, level-1 effect
- BP and WP effects almost always need to be represented by two or more model parameters, using either:
  - *Person-mean-centering* ($WPx_{ti}$ and $PMx_i$): WP ≠ 0?, BP ≠ 0?
  - *Grand-mean-centering* ($TVx_{ti}$ and $PMx_i$): WP ≠ 0?, BP ≠ WP?
  - Both yield equivalent models if the level-1 WP effect is fixed, but not if the level-1 WP effect is random
    - Grand MC → *absolute* effect of $x_{ti}$ varies randomly over people
    - Person MC → *relative* effect of $x_{ti}$ varies randomly over people
    - Use prior theory and empirical data (ML AIC, BIC) to decide

# Time-Varying Predictors in Longitudinal Models

- Topics:
  - Time-varying predictors that fluctuate over time
  - Person-Mean-Centering (PMC)
  - Grand-Mean-Centering (GMC)
  - Model extensions under Person-MC vs. Grand-MC
  - **Time-varying predictors that change over time**



# Baseline Centering for Time-Varying Predictors that Change over Time

- Although using the person mean of the time-varying predictor at level-2 ($PMx_i$) is the most common way to represent the effect of between-person differences, there are other options that sometimes can be more useful

- **Level-2 → X at centering point of time (e.g., $x_{ti}$ at time 0)**
  - Useful if $x_{ti}$ at specific time point conveys useful information, such as baseline level of a covariate in an intervention
  - Useful if $x_{ti}$ is expected to change systematically over time, too

- Create predictors using a variant of PMC → **baseline centering**:
  - Level 1 = $stress_{ti}$ – $\mathbf{stressTime0_i}$ → longitudinal effect
    - L1 represents *change from baseline*, not deviation from own mean
  - Level 2 = $\mathbf{stressTime0_i}$ – C → cross-sectional effect
    - L2 represents effect of *baseline level*, not effect of mean level averaged over time

# Baseline Centering: Caveats

- In using baseline centering instead of person-mean-centering, a complete separation of the BP and WP variance in the time-varying predictor is not obtained:
  - If the time-varying predictor shows change, you are not fitting a model for that change—no separation of true change from error
  - The level-1 predictor for "WP change in X" is both individual differences in change ($U_{1i}$) and residual deviations from change ($e_{ti}$), which should each really have their own relationship to the outcome
  - Therefore, there may be systematic BP differences with regard to the individual slope still contained in the WP change in X predictor (which may be related to BP differences in level at time 0)
- A better option is to use a multivariate model instead, in which a model for change X is fitted for both X and Y
  - Can examine relationships between intercepts, slopes, and residuals as separate model parameters
  - Can be done in MLM programs, but more flexibility in SEM programs



# Multivariate Models via M-SEM

- Person-MC (or baseline centering) is the poor man's version of a model-based decomposition of BP and WP variance, which is necessary when X is treated as a predictor in MLM programs
- Through Multilevel Structural Equation Modeling (M-SEM), it is possible to fit a model for X along with the model for Y
  - It's called SEM because random effects = latent variables, but there is no latent variable measurement model as in traditional uses of SEM
  - Person mean = random intercept variance, WP deviation = residual variance, but can also include random slopes for change over time in X
  - Can directly assess multilevel mediation through simultaneous analysis
  - Some evidence that level-2 effects are less biased (because person mean is not perfectly reliable), but more imprecise (more parameters to estimate)
- What could go wrong? No REML! Good luck fitting interactions!
  - Those involving level-2 effects are modeled as latent variable interactions
  - This requires numeric integration, a very computationally intense way of getting parameter estimates in ML, which may not be possible in all data

# Two-Level Models for Clustered* Data

- Topics:
  - **Fixed vs. random effects for modeling clustered data**
  - **ICC and design effects in clustered data**
  - Group-Mean-Centering vs. Grand-Mean Centering
  - Model extensions under Group-MC and Grand-MC

* *Clustering = Nesting = Grouping…*



# MLM for Clustered Data

- So far we've built models to account for dependency created by repeated measures (time within person)
- Now we examine two-level models for more general examples of nesting/clustering/grouping:
  - Students within schools, athletes within teams
  - Siblings within families, partners within dyads
  - Employees within businesses, patients within doctors
- Residuals of people from same group are likely to be correlated due to group differences (e.g., purposeful grouping or shared experiences create dependency)
- **Recurring theme: You still have to care about group-level variation, even if that's not the point of your study**

# 2 Options for Differences Across Groups

## Represent Group Differences as Fixed Effects

- Include (#groups-1) contrasts for group membership in the **model for the means** (via CLASS)→ so group is NOT another "level"
- Permits inference about differences between specific groups, but you cannot include between-group predictors (group is saturated)
- Snijders & Bosker (1999) ch. 4, p. 44 recommend if #groups < 10ish

## Represent Group Differences as a Random Effect

- Include a **random intercept variance in the model for the variance**, such that group differences become another "level"
- Permits inference about differences across groups more generally, for which you can test effects of between-group predictors
- Better if #groups > 10ish and you want to **predict** group differences



# Empty Means, Random Intercept Model

## MLM for Clustered Data:

- Change in notation:
  - $i$ = level 1, $j$ = level 2
- Level 1:

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

- Level 2:

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

**Residual = person-specific deviation from group's predicted outcome**

**Fixed Intercept =grand mean (because no predictors yet)**

**Random Intercept = group-specific deviation from predicted intercept**

**3** Total Parameters:

**Model for the Means (1):**

- Fixed Intercept $\gamma_{00}$

**Model for the Variance (2):**

- Level-1 Variance of $e_{ij}$ → $\sigma^2_e$
- Level-2 Variance of $U_{0j}$ → $\tau^2_{U_0}$

**Composite equation:**

$$y_{ij} = (\gamma_{00} + U_{0j}) + e_{ij}$$

# Matrices in a Random Intercept Model

**Total predicted data matrix is called V matrix, created from the G [TYPE=UN] and R [TYPE=VC] matrices as follows:**

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^{T} + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}\tau^2_{U_0}\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}\sigma^2_e & 0 & 0 & 0\\ 0 & \sigma^2_e & 0 & 0\\ 0 & 0 & \sigma^2_e & 0\\ 0 & 0 & 0 & \sigma^2_e\end{bmatrix} = \begin{bmatrix}\tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e\end{bmatrix}$$

**VCORR then provides the intraclass correlation, calculated as:**
**ICC = $\tau^2_{U_0}$ / ($\tau^2_{U_0}$ + $\sigma^2_e$)**

$$\begin{bmatrix}1 & ICC & ICC & ICC\\ ICC & 1 & ICC & ICC\\ ICC & ICC & 1 & ICC\\ ICC & ICC & ICC & 1\end{bmatrix}$$

assumes a constant correlation over time

The **G**, **Z**, and **R** matrices still get combined to create the **V** matrix, except that they are now per group. **R** and **V** have *n* rows by *n* columns, in which *n* = # level-1 units, which is now people, not time. Thus, no type of **R** matrix other than VC will be used, and REPEATED is not needed.



# Intraclass Correlation (ICC)

$$ICC = \frac{BG}{BG + WG} = \frac{\text{Intercept Variance}}{\text{Intercept Variance} + \text{Residual Variance}}$$

$$= \frac{\tau^2_{U_0}}{\tau^2_{U_0} + \sigma^2_e}$$

$\tau^2_{U_0}$ → Why don't all groups have the same mean?
$\sigma^2_e$ → Why don't all people from the same group have the same outcome?

- ICC = Proportion of total variance that is between groups
- ICC = Average correlation among persons from same group
- ICC is a standardized way of expressing how much we need to worry about *dependency due to group mean differences*
  **(i.e., ICC is an effect size for *constant* group dependency)**
  - Dependency of other kinds can still be created by differences between groups in the effects of predictors (stay tuned)

# Effects of Clustering on Effective N

- **Design Effect** expresses how much effective sample size needs to be adjusted due to clustering/grouping
- **Design Effect** = ratio of the variance obtained with the given sampling design to the variance obtained for a simple random sample from the same population, given the same total sample size either way

*n* = # level-1 units

- Design Effect = $1 + [(n - 1) * \text{ICC}]$
- Effective sample size → $N_{\text{effective}} = \frac{\#\ \text{Total Observations}}{\text{Design Effect}}$
- As ICC goes UP and cluster size goes UP, the effective sample size goes DOWN
  - See Snijders & Bosker (1999) ch. 3, p. 22-24 for more info



# Design Effects in 2-Level Nesting

- Design Effect = $1 + [(n - 1) * \text{ICC}]$
- Effective sample size → $N_{\text{effective}} = \frac{\#\ \text{Total Observations}}{\text{Design Effect}}$
- *n*=5 patients from each of 100 doctors, ICC = .30?
  - Patients Design Effect = 1 + (4 * .30) = 2.20
  - $\mathbf{N_{effective}}$ = 500 / 2.20 = **227** (not 500)
- *n*=20 students from each of 50 schools, ICC = .05?
  - Students Design Effect = 1 + (19 * .05) = 1.95
  - $\mathbf{N_{effective}}$ = 1000 / 1.95 = **513** (not 1000)

# Does a non-significant ICC mean you can ignore groups and just do a regression?

- Effective sample size depends on BOTH the ICC and the number of people per group: As ICC goes UP and group size goes UP, the effective sample size goes DOWN
  - So there is NO VALUE OF ICC that is "safe" to ignore, not even 0!
  - An ICC=0 in an *empty (unconditional)* model can become ICC>0 after adding level-1 predictors, because reducing the residual variance leads to an increase in the random intercept variance (→ *conditional* ICC > 0)
- So just do a multilevel analysis anyway…
  - Even if "that's not your question"… because people come from groups, you still have to model group dependency appropriately because of:
    - Effect of clustering on level-1 fixed effect SE's → biased SEs
    - Potential for contextual effects of level-1 predictors



# Predictors in MLM for Clustered Data
## Example: Achievement in Students nested in Schools

- Level-2 predictors now refer to Group-Level Variables
  - Can only have fixed or systematically varying effects (level-2 predictors cannot have random effects in a two-level model, same as before)
  - e.g., Does mean school achievement differ b/t rural and urban schools?
- Level-1 predictors now refer to Person-Level Variables
  - Can have fixed, systematically varying, or random effects over groups
  - e.g., Does student achievement differ between boys and girls?
    - Fixed effect: Is there a gender difference in achievement, period?
    - Systematically varying effect: Does the gender effect differ b/t rural and urban schools? (but the gender effect is the same within rural and within urban schools)
    - Random effect: Does the gender effect differ *randomly* across schools?
  - We can skip all the steps for building models for "time" and head straight to predictors (given that level-1 units are exchangeable here)

# Two-Level Models for Clustered* Data

- Topics:
  - Fixed vs. random effects for modeling clustered data
  - ICC and design effects in clustered data
  - **Group-Mean-Centering vs. Grand-Mean Centering**
  - Model extensions under Group-MC and Grand-MC

** Clustering = Nesting = Grouping…*



# Predictors in MLM for Clustered Data

- BUT we still need to distinguish level-2 BG effects from level-1 WG effects of level-1 predictors: NO SMUSHING ALLOWED

- Options for representing level-2 BG variance as a predictor:
  - Use **obtained** group mean of level-1 $x_{ij}$ from your sample (labeled as $\mathbf{GMx_j}$ or $\mathbf{\bar{X}_j}$), centered at a constant so that 0 is a meaningful value
  - Use **actual** group mean of level-1 $x_{ij}$ from outside data (also centered so 0 is meaningful) → better if your sample is not the full population

- Can use either **Group-MC** or **Grand-MC** for level-1 predictors (where Group-MC is like Person-MC in longitudinal models)
  - Level-1 Group-MC → center at a VARIABLE: $\mathbf{WGx_{ij}} = \mathbf{x_{ij}} - \mathbf{\bar{X}_j}$
  - Level-1 Grand-MC → center at a CONSTANT: $\mathbf{L1x_{ij}} = \mathbf{x_{ij}} - \boldsymbol{C}$
    - Use $L1x_{ij}$ when including the actual group mean instead of sample group mean

# 3 Kinds of Effects for Level-1 Predictors

- **Is the Between-Group (BG) effect significant?**
  - Are groups with higher predictor values than other groups also higher on Y than other groups, such that the group mean of the person-level predictor $\mathbf{GMx_j}$ accounts for level-2 random intercept variance ($\tau^2_{U_0}$)?
- **Is the Within-Group (WG) effect significant?**
  - If you have higher predictor values than others in your group, do you also have higher outcomes values than others in your group, such that the within-group deviation $\mathbf{WGx_{ij}}$ accounts for level-1 residual variance ($\sigma^2_e$)?
- **Are the BG and WG effects different sizes: Is there a contextual effect?**
  - After controlling for the absolute value of level-1 predictor for each person, is there still an incremental contribution from having a higher group mean of the predictor (i.e., does a group's general tendency predict $\tau^2_{U_0}$ above and beyond)?
  - If there is no contextual effect, then the BG and WG effects of the level-1 predictor show *convergence*, such that their effects are of equivalent magnitude



# Clustered Data Model with **Group-Mean-Centered Level-1 $x_{ij}$** → WG and BG Effects directly through **separate** parameters

**$x_{ij}$ is group-mean-centered into $WGx_{ij}$, with $GMx_j$ at L2:**

Level 1: $y_{ij} = \beta_{0j} + \beta_{1j}(WGx_{ij}) + e_{ij}$

**$WGx_{ij} = x_{ij} - \bar{X}_j$ → it has only Level-1 WG variation**

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(GMx_j) + U_{0j}$

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

**$GMx_j = \bar{X}_j - C$ → it has only Level-2 BG variation**

**$\gamma_{10}$ = WG main effect of having more $x_{ij}$ than others in your group**

**$\gamma_{01}$ = BG main effect of having more $\bar{X}_j$ than other groups**

**Because $WGx_{ij}$ and $GMx_j$ are uncorrelated, each gets the total effect for its level (WG=L1, BG=L2)**

# 3 Kinds of Effects for Level-1 Predictors

- **What Group-Mean-Centering tells us directly:**

- **Is the Between-Group (BG) effect significant?**
  - Are groups with higher predictor values than other groups also higher on Y than other groups, such that the group mean of the person-level predictor $\mathbf{GMx_j}$ accounts for level-2 random intercept variance ($\tau^2_{U_0}$)?
  - This would be indicated by a significant fixed effect of $\mathbf{GMx_j}$
  - Note: this is NOT controlling for the absolute value of $x_{ij}$ for each person

- **Is the Within-Group (WG) effect significant?**
  - If you have higher predictor values than others in your group, do you also have higher outcomes values than others in your group, such that the within-group deviation $\mathbf{WGx_{ij}}$ accounts for level-1 residual variance ($\sigma^2_e$)?
  - This would be indicated by a significant fixed effect of $\mathbf{WGx_{ij}}$
  - Note: this is represented by the relative value of $x_{ij}$, NOT the absolute value of $x_{ij}$



# 3 Kinds of Effects for Level-1 Predictors

- **What Group-Mean-Centering DOES NOT tell us directly:**

- **Are the BG and WG effects different sizes: Is there a contextual effect?**
  - After controlling for the absolute value of the level-1 predictor for each person, is there still an incremental contribution from the group mean of the predictor (i.e., does a group's general tendency predict $\tau^2_{U_0}$ above and beyond just the person-specific value of the predictor)?
  - In clustered data, the contextual effect is phrased as "after controlling for the individual, what is the additional contribution of the group"?

- **To answer this question about the contextual effect for the incremental contribution of the group mean, we have two options:**
  - Ask for the contextual effect via an ESTIMATE statement in SAS (or TEST in SPSS, or NEW in Mplus, or LINCOM in STATA): `WGx -1 GMx 1`
  - Use **"grand-mean-centering"** for level-1 $x_{ij}$ instead: $\mathbf{L1x_{ij}} = \mathbf{x_{ij}} - \boldsymbol{C}$
    **→ centered at a CONSTANT, NOT A LEVEL-2 VARIABLE**
    - Which constant only matters for what the reference point is; it could be the grand mean or other

# Group-MC vs. Grand-MC for Level-1 Predictors

| Level 2 | | Original | Group-MC Level 1 | Grand-MC Level 1 |
|---|---|---|---|---|
| $\bar{X}_j$ | $GMx_j = \bar{X}_j - 5$ | $x_{ij}$ | $WGx_{ij} = x_{ij} - \bar{X}_j$ | $L1x_{ij} = x_{ij} - 5$ |
| 3 | −2 | 2 | −1 | −3 |
| 3 | −2 | 4 | 1 | −1 |
| 7 | 2 | 6 | −1 | 1 |
| 7 | 2 | 8 | 1 | 3 |

Same $GMx_j$ goes into the model using either way of centering the level-1 variable $x_{ij}$

Using **Group-MC**, $WGx_{ij}$ has NO level-2 BG variation, so it is not correlated with $GMx_j$

Using **Grand-MC**, $L1x_{ij}$ STILL has level-2 BG variation, so it is STILL CORRELATED with $GMx_j$

**So the effects of $GMx_j$ and $L1x_{ij}$ when included together under Grand-MC will be different than their effects would be if they were by themselves…**



# Clustered Data Model with $x_{ij}$ represented at Level 1 Only:

## → WG and BG Effects are **Smushed Together**

**$x_{ij}$ is grand-mean-centered into $L1x_{ij}$, WITHOUT $GMx_j$ at L2:**

Level 1: $y_{ij} = \beta_{0j} + \beta_{1j}(\mathbf{L1x_{ij}}) + e_{ij}$

**$L1x_{ij} = x_{ij} - C$ → it still has both Level-2 BG and Level-1 WG variation**

Level 2: $\beta_{0j} = \gamma_{00} + U_{0j}$

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

**$\gamma_{10}$ = *smushed* WG and BG effects**

**Because $L1x_{ij}$ still contains its original 2 different kinds of variation (BG and WG), its 1 fixed effect has to do the work of 2 predictors!**

**A *smushed* effect is also referred to as the *convergence*, *conflated*, or *composite* effect**

# Convergence (Smushed) Effect of a Level-1 Predictor

$$\text{Convergence Effect: } \gamma_{conv} \approx \frac{\dfrac{\gamma_{BG}}{SE^2_{BG}} + \dfrac{\gamma_{WG}}{SE^2_{WG}}}{\dfrac{1}{SE^2_{BG}} + \dfrac{1}{SE^2_{WG}}}$$

Adapted from Raudenbush & Bryk (2002, p. 138)

- **The convergence effect will often be closer to the within-group effect** (due to larger level-1 sample size and thus smaller SE)
- **It is the rule, not the exception, that between and within effects differ** (Snijders & Bosker, 1999, p. 52-56, and personal experience!)
- However—when grand-mean-centering a level-1 predictor, **convergence is testable** by including a **contextual effect (carried by the group mean)** for how the **BG effect** differs from the **WG effect**…



# Clustered Data Model with **Grand-Mean-Centered Level-1 $x_{ij}$**

→ Model tests difference of WG vs. BG effects (It's been fixed!)

**$x_{ij}$ is grand-mean-centered into $L1x_{ij}$, WITH $GMx_j$ at L2:**

Level 1: $y_{ij} = \beta_{0j} + \beta_{1j}(L1x_{ij}) + e_{ij}$

**$L1x_{ij} = x_{ij} - C$ → it still has both Level-2 BG and Level-1 WG variation**

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(GMx_j) + U_{0j}$

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

**$GMx_j = \bar{X}_j - C$ → it has only Level-2 BG variation**

**$\gamma_{10}$ becomes the WG effect → *unique* level-1 effect after controlling for $GMx_j$**

**$\gamma_{01}$ *becomes* the contextual effect that indicates <u>how the BG effect differs from the WG effect</u> → *unique* level-2 effect after controlling for $L1x_{ij}$ → does group matter beyond individuals?**

# Group-MC and Grand-MC Models are Equivalent Given a Fixed Level-1 Main Effect Only

**Group-MC:** $WGx_{ij} = x_{ij} - GMx_j$

Level-1: $y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - GMx_j) + e_{ij}$

Level-2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(GMx_j) + U_{0j}$

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

$\rightarrow y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(x_{ij} - GMx_j) + U_{0j} + e_{ij}$

$\rightarrow y_{ij} = \gamma_{00} + (\gamma_{01} - \gamma_{10})(GMx_j) + \gamma_{10}(x_{ij}) + U_{0j} + e_{ij}$

**Composite Model:**
← As Group-MC
← As Grand-MC

**Grand-MC:** $L1x_{ij} = x_{ij}$

Level-1: $y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij}) + e_{ij}$

Level-2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(GMx_j) + U_{0j}$

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

$\rightarrow y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(x_{ij}) + U_{0j} + e_{ij}$

| Effect | Group-MC | Grand-MC |
|---|---|---|
| Intercept | $\gamma_{00}$ | $\gamma_{00}$ |
| WG Effect | $\gamma_{10}$ | $\gamma_{10}$ |
| Contextual | $\gamma_{01} - \gamma_{10}$ | $\gamma_{01}$ |
| BG Effect | $\gamma_{01}$ | $\gamma_{01} + \gamma_{10}$ |



# Contextual Effects in Clustered Data

- Group-MC is equivalent to Grand-MC if the group mean of the level-1 predictor is included and the level-1 effect is not random
- Grand-MC may be more convenient in clustered data due to its ability to directly provide contextual effects

- Example: Effect of SES for students (nested in schools) on achievement:

- **Group-MC** of level-1 student $SES_{ij}$, school mean $\overline{SES}_j$ included at level 2
  - Level-1 **WG** effect: Effect of being rich kid relative to your school (is already purely WG because of centering around $\overline{SES}_j$)
  - Level-2 **BG** effect: Effect of going to a rich school NOT controlling for kid $SES_{ij}$

- **Grand-MC** of level-1 student $SES_{ij}$, school mean $\overline{SES}_j$ included at level 2
  - Level-1 **WG** effect: Effect of being rich kid relative to your school (is purely WG after *statistically* controlling for $\overline{SES}_j$)
  - Level-2 **Contextual** effect: Incremental effect of going to a rich school (after *statistically* controlling for student SES)

# 3 Kinds of Effects for Level-1 Predictors

- **Is the Between-Group (BG) effect significant?**
  - Are groups with higher predictor values <u>than other groups</u> also higher on Y <u>than other groups</u>, such that the group mean of the person-level predictor $\mathbf{GMx_j}$ accounts for level-2 random intercept variance ($\tau^2_{U_0}$)?
  - Given directly by level-2 effect of $GMx_j$ if using Group-MC for the level-1 predictor (or can be requested via ESTIMATE if using Grand-MC for the level-1 predictor)

- **Is the Within-Group (WG) effect significant?**
  - If you have higher predictor values <u>than others in your group</u>, do you also have higher outcomes values <u>than others in your group</u>, such that the within-group deviation $\mathbf{WGx_{ij}}$ accounts for level-1 residual variance ($\sigma^2_e$)?
  - Given directly by the level-1 effect of $WGx_{ij}$ if using Group-MC —*OR* — given directly by the level-1 effect of $L1x_{ij}$ if using Grand-MC and including $GMx_j$ at level 2 (without $GMx_j$, the level-1 effect of $L1x_{ij}$ if using Grand-MC is the smushed effect)

- **Are the BG and WG effects different sizes: Is there a contextual effect?**
  - After controlling for the absolute value of the level-1 predictor for each person, is there still <u>an incremental contribution from the group mean of the predictor</u> (i.e., does a group's general tendency predict $\tau^2_{U_0}$ above and beyond the person-specific predictor value)?
  - Given directly by level-2 effect of $GMx_j$ if using Grand-MC for the level-1 predictor (or can be requested via ESTIMATE if using Group-MC for the level-1 predictor)



# Variance Accounted For By Level-2 Predictors

- **Fixed effects of level 2 predictors *by themselves*:**
  - Level-2 (BG) main effects reduce level-2 (BG) random intercept variance
  - Level-2 (BG) interactions also reduce level-2 (BG) random intercept variance

- **Fixed effects of *cross-level interactions* (level 1* level 2):**
  - If the interacting level-1 predictor is <u>random</u>, any cross-level interaction with it will reduce its corresponding level-2 BG random slope variance (that line's U)
  - If the interacting level-1 predictor <u>not random</u>, any cross-level interaction with it will reduce the level-1 WG residual variance instead
    - This is because the level-2 BG random slope variance would have been created by decomposing the level-1 residual variance in the first place
    - The level-1 effect would then be called "**systematically varying**" to reflect a compromise between "fixed" (all the same) and "random" (all different)—it's not that each group needs their own slope, but that the slope varies systematically across groups as a function of a known group predictor (and not otherwise)

# Variance Accounted For By Level-1 Predictors

- **Fixed effects of level 1 predictors *by themselves*:**
  - Level-1 (WG) main effects reduce Level-1 (WG) residual variance
  - Level-1 (WG) interactions also reduce Level-1 (WG) residual variance

- **What happens at level 2 depends on what kind of variance the level-1 predictor has:**
  - If the level-1 predictor ALSO has level-2 variance (e.g., Grand-MC predictors), then its level-2 variance will also likely reduce level-2 random intercept variance
  - If the level-1 predictor DOES NOT have level-2 variance (e.g., Group-MC predictors), then its reduction in the level-1 residual variance will cause an INCREASE in level-2 random intercept variance
    - Same thing happens with Grand-MC level-1 predictors, but you don't generally see it
  - It's just an artifact that the estimate of true random intercept variance is:

    True $\tau^2_{U_0}$ = observed $\tau^2_{U_0} - \frac{\sigma^2_e}{n}$ → so if only $\sigma^2_e$ decreases, $\tau^2_{U_0}$ increases



# Two-Level Models for Clustered* Data

- Topics:
  - Fixed vs. random effects for modeling clustered data
  - ICC and design effects in clustered data
  - Group-Mean-Centering vs. Grand-Mean Centering
  - **Model extensions under Group-MC and Grand-MC**

** Clustering = Nesting = Grouping…*

# The Joy of Interactions Involving Level-1 Predictors

- <u>Must consider interactions with both its BG and WG parts:</u>
- Example: Does the effect of employee motivation ($x_{ij}$) on employee performance interact with type of business (for profit or non-profit; $Type_j$)?

- <u>Group-Mean-Centering:</u>
  - $WGx_{ij} * Type_j$ → Does the WG motivation effect differ between business types?
  - $GMx_j * Type_j$ → Does the BG motivation effect differ between business types?
    - Moderation of <u>total</u> group motivation effect (not controlling for individual motivation)
    - If forgotten, then $Type_j$ moderates the motivation effect only at level 1 (WG, not BG)

- <u>Grand-Mean-Centering:</u>
  - $L1x_{ij} * Type_j$ → Does the WG motivation effect differ between business types?
  - $GMx_j * Type_j$ → Does the *contextual* motivation effect differ b/t business types?
    - Moderation of <u>incremental</u> group motivation effect <u>controlling for employee motivation</u> (moderation of the "boost" in group performance from working with motivated people)
    - If forgotten, then although the level-1 main effect of motivation has been un-smushed via the main effect of $GMx_j$, the interaction of $L1x_{ij} * Type_j$ would still be smushed



# Interactions with Level-1 Predictors: Example: Employee Motivation ($x_{ij}$) by Business Type ($Type_j$)

**<u>Group-MC:</u>** $WGx_{ij} = x_{ij} - GMx_j$

**Level-1:** $y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - GMx_j) + e_{ij}$

**Level-2:** $\beta_{0j} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{02}(Type_j) + \gamma_{03}(Type_j)(GMx_j) + U_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}(Sex_i)$

**Composite:** $y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(x_{ij} - GMx_j) + U_{0j} + e_{ij}$
$+ \gamma_{02}(Type_j) + \gamma_{03}(Type_j)(GMx_j) + \gamma_{11}(Type_j)(x_{ij} - GMx_j)$

---

**<u>Grand-MC:</u>** $L1x_{ij} = x_{ij}$

**Level-1:** $y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij}) + e_{ij}$

**Level-2:** $\beta_{0j} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{02}(Type_j) + \gamma_{03}(Type_j)(GMx_j) + U_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}(Type_j)$

**Composite:** $y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(x_{ij}) + U_{0j} + e_{ij}$
$+ \gamma_{02}(Type_j) + \gamma_{03}(Type_j)(GMx_j) + \gamma_{11}(Type_j)(x_{ij})$

# Interactions Involving Level-1 Predictors Belong at Both Levels of the Model

**On the left below → Group-MC:** $WGx_{ij} = x_{ij} - GMx_j$

$y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(x_{ij} - GMx_j) + U_{0j} + e_{ij}$
$+ \gamma_{02}(Type_j) + \gamma_{03}(Type_j)(GMx_j) + \gamma_{11}(Type_j)(x_{ij} - GMx_j)$ ← **As Group-MC**

$y_{ij} = \gamma_{00} + (\gamma_{01} - \gamma_{10})(GMx_j) + \gamma_{10}(x_{ij}) + U_{0j} + e_{ij}$
$+ \gamma_{02}(Type_j) + (\gamma_{03} - \gamma_{11})(Type_j)(GMx_j) + \gamma_{11}(Type_j)(x_{ij})$ ← **As Grand-MC**

---

**On the right below → Grand-MC:** $L1x_{ij} = x_{ij}$

$y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(x_{ij}) + U_{0j} + e_{ij}$
$+ \gamma_{02}(Type_j) + \gamma_{03}(Type_j)(GMx_j) + \gamma_{11}(Type_j)(x_{ij})$

After adding an interaction for $Type_j$ with $x_{ij}$ at both levels, then the Group-MC and Grand-MC models are equivalent

---

**Intercept:** $\gamma_{00} = \gamma_{00}$ **BG Effect:** $\gamma_{01} = \gamma_{01} + \gamma_{10}$ **Contextual:** $\gamma_{01} = \gamma_{01} - \gamma_{10}$

**WG Effect:** $\gamma_{10} = \gamma_{10}$ **BG*Type Effect:** $\gamma_{03} = \gamma_{03} + \gamma_{11}$ **Contextual*Type:** $\gamma_{03} = \gamma_{03} - \gamma_{11}$

**Type Effect:** $\gamma_{20} = \gamma_{20}$ **BG*WG or Contextual*WG is the same:** $\gamma_{11} = \gamma_{11}$



# Intra-variable Interactions

- Still must consider interactions with both its BG and WG parts!
- Example: Does the effect of employee motivation ($x_{ij}$) on employee performance interact with business group mean motivation ($GMx_j$)?
- Group-Mean-Centering:
  - $WGx_{ij} * GMx_j$ → Does the WG motivation effect differ by group motivation?
  - $GMx_j * GMx_j$ → Does the BG motivation effect differ by group motivation?
    - Moderation of total group motivation effect (not controlling for individual motivation)
    - If forgotten, then $GMx_j$ moderates the motivation effect only at level 1 (WG, not BG)
- Grand-Mean-Centering:
  - $L1x_{ij} * GMx_j$ → Does the WG motivation effect differ by group motivation?
  - $GMx_j * GMx_j$ → Does the *contextual* motivation effect differ by group motiv.?
    - Moderation of incremental group motivation effect controlling for employee motivation (moderation of the boost in group performance from working with motivated people)
    - If forgotten, then although the level-1 main effect of motivation has been un-smushed via the main effect of $GMx_j$, the interaction of $L1x_{ij} * GMx_j$ would still be smushed

# Intra-variable Interactions:
## Example: Employee Motivation ($x_{ij}$) by Business Mean Motivation ($GMx_j$)

**Group-MC:** $WGx_{ij} = x_{ij} - GMx_j$

**Level-1:** $y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - GMx_j) + e_{ij}$

**Level-2:** $\beta_{0j} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{02}(GMx_j)(GMx_j) + U_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}(GMx_j)$

**Composite:** $y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(x_{ij} - GMx_j) + U_{0j} + e_{ij}$
$+ \gamma_{02}(GMx_j)(GMx_j) + \gamma_{11}(GMx_j)(x_{ij} - GMx_j)$

---

**Grand-MC:** $L1x_{ij} = x_{ij}$

**Level-1:** $y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij}) + e_{ij}$

**Level-2:** $\beta_{0j} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{02}(GMx_j)(GMx_j) + U_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}(GMx_j)$

**Composite:** $y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(x_{ij}) + U_{0j} + e_{ij}$
$+ \gamma_{02}(GMx_j)(GMx_j) + \gamma_{11}(GMx_j)(x_{ij})$



# Intra-variable Interactions:
## Example: Employee Motivation ($x_{ij}$) by Business Mean Motivation ($GMx_j$)

**On the left below → Group-MC:** $WGx_{ij} = x_{ij} - GMx_j$

$y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(x_{ij} - GMx_j) + U_{0j} + e_{ij}$
$+ \gamma_{02}(GMx_j)(GMx_j) + \gamma_{11}(GMx_j)(x_{ij} - GMx_j)$

← As Group-MC

$y_{ij} = \gamma_{00} + (\gamma_{01} - \gamma_{10})(GMx_j) + \gamma_{10}(x_{ij}) + U_{0j} + e_{ij}$
$+ (\gamma_{02} - \gamma_{11})(GMx_j)(GMx_j) + \gamma_{11}(GMx_j)(x_{ij})$

← As Grand-MC

---

**On the right below → Grand-MC:** $L1x_{ij} = x_{ij}$

$y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(x_{ij}) + U_{0j} + e_{ij}$
$+ \gamma_{02}(GMx_j)(GMx_j) + \gamma_{11}(GMx_j)(x_{ij})$

After adding an interaction for $Type_j$ with $x_{ij}$ at both levels, then the Group-MC and Grand-MC models are equivalent

---

**Intercept:** $\gamma_{00} = \gamma_{00}$ **BG Effect:** $\gamma_{01} = \gamma_{01} + \gamma_{10}$ **Contextual:** $\gamma_{01} = \gamma_{01} - \gamma_{10}$

**WG Effect:** $\gamma_{10} = \gamma_{10}$ **BG² Effect:** $\gamma_{02} = \gamma_{02} + \gamma_{11}$ **Contextual²:** $\gamma_{02} = \gamma_{02} - \gamma_{11}$

**BG*WG or Contextual*WG is the same:** $\gamma_{11} = \gamma_{11}$

# When Group-MC ≠ Grand-MC: Random Effects of Level-1 Predictors

**Group-MC:** $WGx_{ij} = x_{ij} - GMx_j$

Level-1: $y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij} - GMx_j) + e_{ij}$

Level-2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(GMx_j) + U_{0j}$

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

Variance due to $GMx_j$ is removed from the random slope in Group-MC.

$\rightarrow y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(x_{ij} - GMx_j) + U_{0j} + U_{1j}(x_{ij} - GMx_j) + e_{ij}$

**Grand-MC:** $L1x_{ij} = x_{ij}$

Level-1: $y_{ij} = \beta_{0j} + \beta_{1j}(x_{ij}) + e_{ij}$

Level-2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(GMx_j) + U_{0j}$

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

Variance due to $GMx_j$ is still part of the random slope in Grand-MC. So these models cannot be made equivalent.

$\rightarrow y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(x_{ij}) + U_{0j} + U_{1j}(x_{ij}) + e_{ij}$

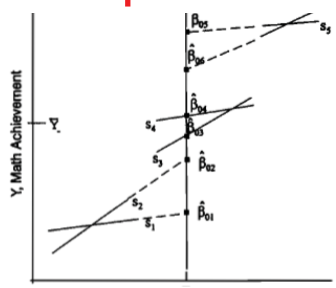


# Random Effects of Level-1 Predictors

- **Random intercepts** mean different things under each model:
  - **Group-MC** → Group differences at **$WGx_{ij}$ =0** (that every group has)
  - **Grand-MC** → Group differences at **$L1x_{ij}$=0** (that not every group will have)
- **Differential shrinkage of the random intercepts** results from differential reliability of the intercept data across models:
  - Group-MC → Won't affect shrinkage of slopes unless highly correlated
  - Grand-MC → Will affect shrinkage of slopes due to forced extrapolation
- As a result, the **random slope variance may be smaller** under Grand-MC than under Group-MC
  - Problem worsens with greater ICC of level-1 predictor (more extrapolation)
  - Anecdotal example was presented in Raudenbush & Bryk (2002; chapter 5)

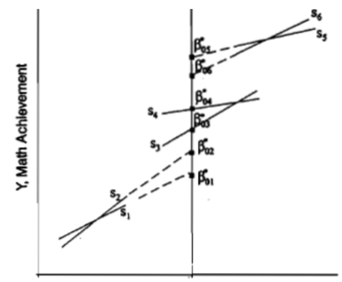
# Bias in Random Slope Variance

**OLS Per-Group Estimates**




**Level-1 X**

**EB Shrunken Estimates**




**Level-1 X**

Top right: Intercepts and slopes are homogenized in Grand-MC because of intercept extrapolation

Bottom: Downwardly-biased random slope variance in Grand-MC relative to Group-MC

| *Unconditional Results* | *Conditional Results* |
|---|---|
| **Group-MC** | |
| $\hat{\mathbf{T}} = \begin{bmatrix} 8.68 & 0.05 \\ 0.05 & 0.68 \end{bmatrix}$ $\hat{\sigma}^2 = 36.70$ | $\hat{\mathbf{T}} = \begin{bmatrix} 2.38 & 0.19 \\ 0.19 & \boxed{0.15} \end{bmatrix}$ $\hat{\sigma}^2 = 36.70$ |
| **Grand-MC** | |
| $\hat{\mathbf{T}} = \begin{bmatrix} 4.83 & -0.15 \\ -0.15 & 0.42 \end{bmatrix}$ $\hat{\sigma}^2 = 36.83$ | $\hat{\mathbf{T}} = \begin{bmatrix} 2.41 & 0.19 \\ 0.19 & \boxed{0.06} \end{bmatrix}$ $\hat{\sigma}^2 = 36.74$ |



# MLM for Clustered Data: Summary

- Models now come in only two kinds: “empty” and “conditional”
  - The lack of a comparable dimension to “time” simplifies things greatly!
- L2 = Between-Group, L1 = Within-Group (between-person)
  - Level-2 predictors are group variables: can have fixed or systematically varying effects (but not random effects in two-level models)
  - Level-1 predictors are person variables: can have fixed, random, or systematically varying effects
- No smushing main effects or interactions of level-1 predictors:
  - Group-MC at Level 1: Get L1=WG and L2=BG effects directly
  - Grand-MC at Level 1: Get L1=WG and L2=contextual effects directly
    - As long as some representation of the L1 effect is included in L2; otherwise, the L1 effect (and any interactions thereof) will be smushed

# Three-Level Models for Clustered Longitudinal Data

- Topics:
  - **Decomposing variation across three levels in clustered longitudinal data**
  - Unconditional (time only) model specification
  - Conditional (other predictors) model specification
  - Other kinds of three-level designs



# What determines the number of levels?

- **Answer: the model for the outcome variance ONLY**
- How many dimensions of sampling in the outcome?
  - Time within person → 2-level model
  - Time within person within family → 3-level model
  - Time within person within family within country → 4-level model
  - Sampling dimensions may also be crossed instead of nested, or may be modeled with fixed effects if the # units is small
- Need at least one pile of variance per dimension (for 3 levels, that's 2 sets of random effects and a residual)
  - Include whatever predictors you want for each level, but keep in mind that the usefulness of your predictors will be constrained by how much Y variance exists in its relevant sampling dimension

# Empty Means, 3-Level Random Intercept Model

Notation: $t$ = level-1 time, $i$ = level-2 person, $j$ = level-3 group

**Level 1:** $y_{tij} = \beta_{0ij} + e_{tij}$

**Residual** = time-specific deviation from **person's** predicted outcome

**Level 2:** $\beta_{0ij} = \delta_{00j} + U_{0ij}$

**Person Random Intercept** = person-specific deviation from **group's** predicted outcome

**Level 3:** $\delta_{00j} = \gamma_{000} + V_{00j}$

**Fixed Intercept** =grand mean (because no predictors yet)

**Group Random Intercept** = group-specific deviation from fixed intercept

**3** Total Parameters:

**Model for the Means (1):**

- Fixed Intercept $\gamma_{00}$

**Model for the Variance (2):**

- Level-1 Variance of $e_{tij}$ → $\sigma_e^2$
- Level-2 Variance of $U_{0ij}$ → $\tau_{U_0}^2$
- Level-3 Variance of $V_{00j}$ → $\tau_{V_{00}}^2$

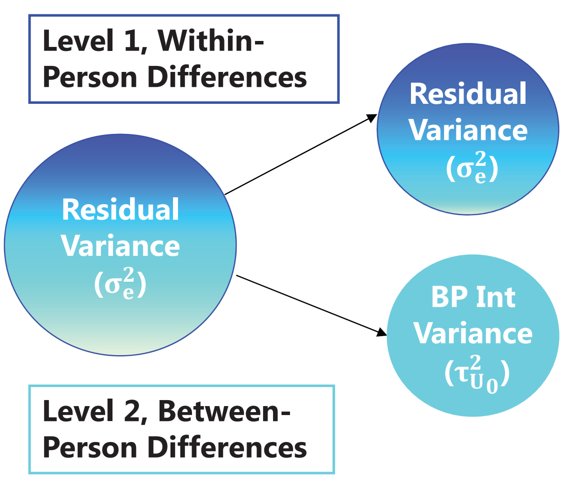
**Composite equation:**

$y_{tij} = \gamma_{000} + V_{00j} + U_{0ij} + e_{tij}$



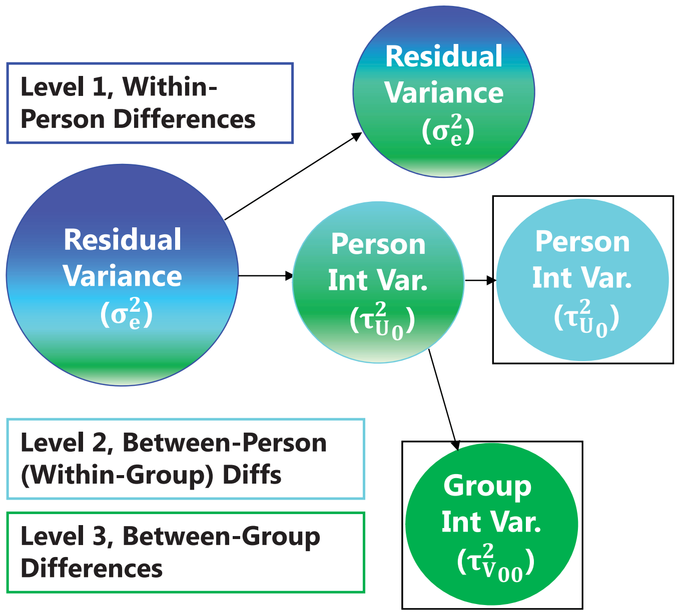
# 2-Level Random Intercept Model

- Where does each kind of person dependency go? Into a new random effects variance component (or "pile" of variance):
- Let's start with an empty means, random intercept 2-level model for time within person:



# 3-Level Random Intercept Model

- Now let's see what happens in an empty means, random intercept 3-level model of time within person within groups:






# ICCs in a 3-Level Random Intercept Model
## Example: Time within Person within Group

- **ICC for level 2 (and level 3) relative to level 1:**

- $\text{ICC}_{\text{L2}} = \frac{\text{Between-Person}}{\text{Total}} = \frac{\text{L3+L2}}{\text{L3+L2+L1}} = \frac{\tau_{V_{00}}^2 + \tau_{U_0}^2}{\tau_{V_{00}}^2 + \tau_{U_0}^2 + \sigma_e^2}$

  → This ICC expresses similarity of occasions from same person (and by definition, from the same group) → of the ***total variation in Y***, how much of it is ***between persons, or not due to time***?

- **ICC for level 3 relative to level 2 (ignoring level 1):**

- $\text{ICC}_{\text{L3}} = \frac{\text{Between-Group}}{\text{Between-Person}} = \frac{\text{L3}}{\text{L3+L2}} = \frac{\tau_{V_{00}}^2}{\tau_{V_{00}}^2 + \tau_{U_0}^2}$

  → This ICC expresses similarity of persons from same group (ignoring within-person variation over time) → of ***that total between-person variation in Y***, how much of that is actually ***between groups***?

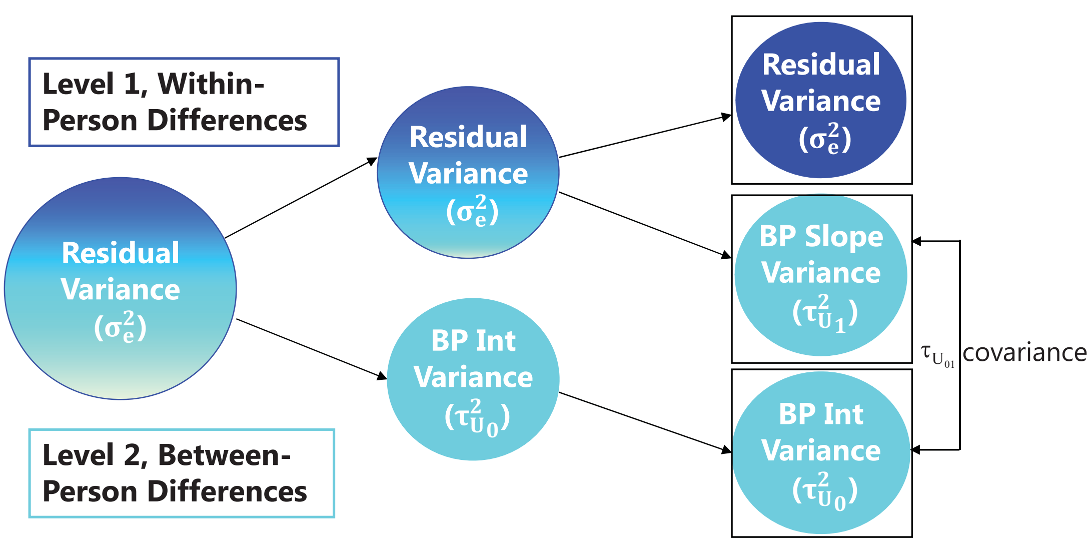
# Three-Level Models for Clustered Longitudinal Data

- Topics:
  - Decomposing variation across three levels in clustered longitudinal data
  - **Unconditional (time only) model specification**
  - Conditional (other predictors) model specification
  - Other kinds of three-level designs



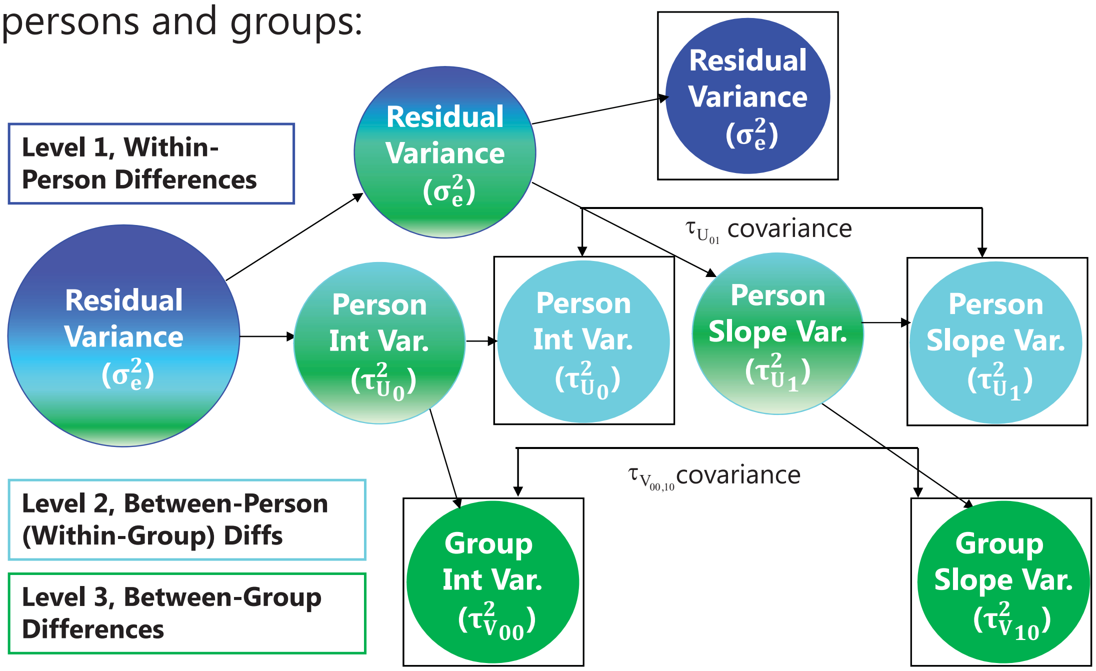
# 2-Level Random Slope Model

- What about time? After adding fixed effects of time, we can add random effects of time over persons in a 2-level model:



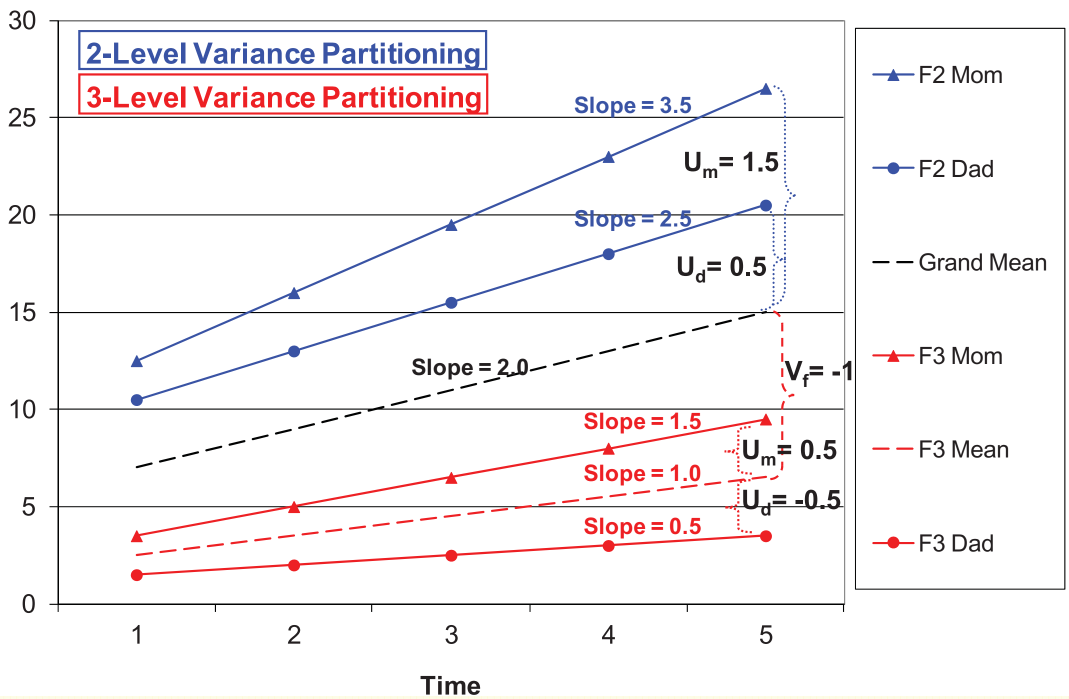
# 3-Level Random Slope Model

- In a 3-level model, we can have random effects of time over persons and groups:






# Random Time Slopes at both Level 2 AND Level 3? An example with family as group:



# 3-Level Random Time Slope Model

Notation: $t$ = level-1 time, $i$ = level-2 person, $j$ = level-3 group

**Level 1:** $y_{tij} = \beta_{0ij} + \beta_{1ij}(Time_{tij}) + e_{tij}$

**Residual** = time-specific deviation from **person's** predicted growth line ($\sigma_e^2$)

**Level 2:** $\beta_{0ij} = \delta_{00j} + U_{0ij}$

$\beta_{1ij} = \delta_{10j} + U_{1ij}$

**Person Random Intercept and Slope** = person-specific deviations from **group's** predicted intercept, slope ($\tau_{U_0}^2$, $\tau_{U_1}^2$, $\tau_{U_{01}}$)

**Level 3:** $\delta_{00j} = \gamma_{000} + V_{00j}$

$\delta_{10j} = \gamma_{100} + V_{10j}$

**Group Random Intercept and Slope** = group-specific deviations from fixed intercept, slope ($\tau_{V_{00}}^2$, $\tau_{V_{10}}^2$, $\tau_{V_{00,10}}$)

**Fixed Intercept, Fixed Linear Time Slope**

**Composite equation (9 parameters):**

$$y_{tij} = (\gamma_{000} + V_{00j} + U_{0ij}) + (\gamma_{100} + V_{10j} + U_{1ij})(Time_{tij}) + e_{tij}$$



# ICCs for Random Intercepts and Slopes

- Once random slopes are included at both level-3 and level-2, ICCs can be computed for the random intercepts and slopes specifically (which would be the level-3 type of ICC)

$$ICC_{Int} = \frac{\text{Between} - \text{Group}}{\text{Between} - \text{Person}} = \frac{\text{L3 Int}}{\text{L3 Int} + \text{L2 Int}} = \frac{\tau_{V_{00}}^2}{\tau_{V_{00}}^2 + \tau_{U_0}^2}$$

$$ICC_{Slope} = \frac{\text{Between} - \text{Group}}{\text{Between} - \text{Person}} = \frac{\text{L3 Slope}}{\text{L3 Slope} + \text{L2 Slope}} = \frac{\tau_{V_{10}}^2}{\tau_{V_{10}}^2 + \tau_{U_1}^2}$$

- Can be computed for any level-1 slope that is random at both levels (e.g., linear and quadratic time, time-varying predictors)
- Be careful when the model is uneven across levels, though

$$\frac{\text{Random Level 2: int, linear, quad}}{\text{Random Level 3: int, linear}} \rightarrow \frac{\text{Linear is when time} = 0}{\text{Linear is at any occasion}}$$

## More on Random Slopes in 3-Level Models

- Any level-1 predictor can have a random effect over level 2, level 3, or over both levels, but I recommend working your way **UP the higher levels** for assessing random effects…
  - e.g., Does the effect of time vary over persons?
  - If so, does the effect of time vary over groups, too? → Is there a commonality in how people from the same group change over time?
- … because random effects at level 3 only are possible but unlikely (e.g., means everyone in the group changes the same)
- Level-2 predictors can also have random effects over level 3
  - e.g., Does the effect of a person characteristic vary over groups?
- Level-1, level-2, and level-1 by level-2 cross-level interactions can all have random effects over level 3, too
  - But tread carefully! The more random effects you have, the more likely you are to have convergence problems ("not positive definite")



# Three-Level Models for Clustered Longitudinal Data

- Topics:
  - Decomposing variation across three levels in clustered longitudinal data
  - Unconditional (time only) model specification
  - **Conditional (other predictors) model specification**
  - Other kinds of three-level designs

# Conditional Model Specification

- Remember separating between- and within-person effects?
  Now there are 3 potential effects for any level-1 predictor!
  - Example: Effect of stress on wellbeing, both measured over time within person within families:
  - **Level 1** (Time): During **Times** of more stress, people have lower (time-specific) wellbeing than in times of less stress
  - **Level 2** (Person): **People** in the family who have more stress have lower (person average) wellbeing than people in the family who have less stress
  - **Level 3** (Family): **Families** who have more stress have lower (family average) wellbeing than families who have less stress

- 2 potential effects for any level-2 predictor, also
  - Example: Effect of baseline level of person coping skills in same design:
  - **Level 2** (Person): **People** in the family who cope better have better (person average) wellbeing than people in the family who cope worse
  - **Level 3** (Family): **Families** who cope better have better (family average) wellbeing than families who cope worse



# Separate Total Effects Per Level Using Person/Group-Mean-Centering

- **Level 1 (Time):** *Time-varying stress relative to person mean*
  → $WPstress_{tij}$ = $Stress_{tij}$ – $PersonMeanStress_{ij}$
  → Direct tests if within-person effect ≠ 0?
  → **Total** within-person effect of having more stress ***than usual*** ≠ 0?

- **Level 2 (Person):** *Person mean stress relative to family*
  → $WFstress_{ij}$ = $PersonMeanStress_{ij}$ – $FamilyMeanStress_{j}$
  → Direct tests if within-family effect ≠ 0?
  → **Total** effect of having more stress ***than other family members*** ≠ 0?

- **Level 3 (Family):** *Family mean stress relative to all families (from constant)*
  → $BFstress_{j}$ = $FamilyMeanStress_{j}$ – C
  → Direct tests if between-family effect ≠ 0?
  → **Total** effect of having more stress ***than other families*** ≠ 0?

# Separate Total Effects Per Level Using Person/Group-Mean-Centering

Notation: $t$ = level-1 time, $i$ = level-2 person, $j$ = level-3 group
PM = person mean, FM = family mean, C = centering constant

**Level 1:** $y_{tij} = \beta_{0ij} + \beta_{1ij}(Time_{tij}) + \beta_{2ij}(Stress_{tij} - PMstress_{ij}) + e_{tij}$

**Level 2:** $\beta_{0ij} = \delta_{00j} + \delta_{01j}(PMstress_{ij} - FMstress_{j}) + U_{0ij}$

$\beta_{1ij} = \delta_{10j} + U_{1ij}$

$\beta_{2ij} = \delta_{20j} + (U_{2ij})$

**Level 3:** $\delta_{00j} = \gamma_{000} + \gamma_{001}(FMstress_{j} - C) + V_{00j}$ — Fixed intercept, Between-family stress main effect

$\delta_{01j} = \gamma_{010} + (V_{01j})$ — Within-family stress main effect

$\delta_{10j} = \gamma_{100} + V_{10j}$ — Time main effect

$\delta_{20j} = \gamma_{200} + (V_{20j})$ — Within-person stress main effect



# Contextual Effects Per Level Using Grand-Mean-Centering

- **Level 1 (Time):** *Time-varying stress (relative to sample constant)*
  - → $TVstress_{tij}$ = $Stress_{tij}$ – C
  - → Direct tests if within-person effect ≠ 0?
  - → **Total** within-person effect of having more stress ***than usual*** ≠ 0?
- **Level 2 (Person):** *Person mean stress (relative to sample constant)*
  - → $BPstress_{ij}$ = $PersonMeanStress_{ij}$ – C
  - → Direct tests if within-person and within-family effects ≠ ?
  - → **Contextual** effect of having more stress ***than other family members*** ≠ 0?
- **Level 3 (Family):** *Family mean stress relative to all families (from constant)*
  - → $BFstress_{j}$ = $FamilyMeanStress_{j}$ – C
  - → Direct tests if within-family and between-family effects ≠ ?
  - → **Contextual** effect of having more stress ***than other families*** ≠ 0?

# Contextual Effects Per Level Using Grand-Mean-Centering

Notation: $t$ = level-1 time, $i$ = level-2 person, $j$ = level-3 group
PM = person mean, FM = family mean, C = centering constant

**Level 1:** $y_{tij} = \beta_{0ij} + \beta_{1ij}(Time_{tij}) + \beta_{2ij}(Stress_{tij} - C) + e_{tij}$

**Level 2:** $\beta_{0ij} = \delta_{00j} + \delta_{01j}(PMstress_{ij} - C) + U_{0ij}$

$\beta_{1ij} = \delta_{10j} + U_{1ij}$

$\beta_{2ij} = \delta_{20j} + (U_{2ij})$

**Level 3:** $\delta_{00j} = \gamma_{000} + \gamma_{001}(FMstress_{j} - C) + V_{00j}$ — **Fixed intercept, Contextual family stress main effect**

$\delta_{01j} = \gamma_{010} + (V_{01j})$ — **Contextual within-family stress main effect**

$\delta_{10j} = \gamma_{100} + V_{10j}$ — **Time main effect**

$\delta_{20j} = \gamma_{200} + (V_{20j})$ — **Within-person stress main effect**



# What does it mean to omit higher-level effects under each centering method?

- **Person-MC**: Removing terms means the effect at that level does not exist (= 0)
  - Remove L3 effect? Assume L3 Between-Family effect = 0
    - *L1 effect = Within-Person effect, L2 effect = Within-Family effect*
  - Then remove L2 effect? Assume L2 Within-Family effect = 0
    - *L1 effect = Within-Person effect*
- **Grand-MC**: Removing terms means the effect at that level is equivalent to the effect at the level beneath it
  - Remove L3 effect? Assume L3 Between-Family = L2 Within-Family effect
    - *L1 effect = Within-Person effect, L2 effect = 'smushed' WF and BF effects*
  - Then remove L2 effect? Assume L2 Between-Person effect = L1 effect
    - *L1 'smushed' = Within-Person, Within-Family, and Between-Family effects*

# Interactions belong at each level, too…

- Example: Is the effect of stress on wellbeing moderated by time-invariant person coping? Using person/group-MC…
- **Stress Effects**
  - **Level 1 (Time):** $WPstress_{tij}$ = $Stress_{tij}$ – $PersonMeanStress_{ij}$
  - **Level 2 (Person):** $WFstress_{ij}$ = $PersonMeanStress_{ij}$ – $FamilyMeanStress_{j}$
  - **Level 3 (Family):** $BFstress_{j}$ = $FamilyMeanStress_{j}$ – C
- **Coping Effects**
  - **Level 2 (Person):** $WFcope_{ij}$ = $Cope_{ij}$ – $FamilyMeanCope_{j}$
  - **Level 3 (Family):** $BFcope_{j}$ = $FamilyMeanCope_{j}$ – C
- **Interaction Effects**
- With level 1 stress: $WPstress_{tij}$ * $WFcope_{ij}$, $WPstress_{tij}$ * $BFcope_{j}$
- With level 2 stress: $WFstress_{ij}$ * $WFcope_{ij}$, ($WFstress_{ij}$ * $BFcope_{j}$)
- With level 3 stress: $BFstress_{j}$ * $BFcope_{j}$, ($BFstress_{j}$ * $WFcope_{ij}$)



# Interactions belong at each level, too…

Notation: $t$ = level-1 time, $i$ = level-2 person, $j$ = level-3 group
PM = person mean, FM = family mean, C = centering constant

**Level 1:** $$y_{tij} = \beta_{0ij} + \beta_{1ij}(Time_{tij}) + \beta_{2ij}(Stress_{tij} - PMstress_{ij}) + e_{tij}$$

**Level 2:**
$$\beta_{0ij} = \delta_{00j} + \delta_{01j}(PMstress_{ij} - FMstress_{j}) + \delta_{02j}(Cope_{ij} - FMcope_{j}) + \delta_{03j}(PMstress_{ij} - FMstress_{j})(Cope_{ij} - FMcope_{j}) + U_{0ij}$$
$$\beta_{1ij} = \delta_{10j} + U_{1ij}$$
$$\beta_{2ij} = \delta_{20j} + \delta_{21j}(Cope_{ij} - FMcope_{j}) + (U_{2ij})$$

**Level 3:**
$$\delta_{00j} = \gamma_{000} + \gamma_{001}(FMstress_{j} - C) + \gamma_{002}(FMcope_{j} - C) + \gamma_{003}(FMstress_{j} - C)(FMcope_{j} - C) + V_{00j}$$
$$\delta_{01j} = \gamma_{010} + (V_{01j}) \quad \delta_{02j} = \gamma_{020} + (V_{02j}) \quad \delta_{03j} = \gamma_{030} + (V_{03j})$$
$$\delta_{10j} = \gamma_{100} + V_{10j}$$
$$\delta_{20j} = \gamma_{200} + \gamma_{202}(FMcope_{j} - C) + (V_{20j}) \quad \delta_{21j} = \gamma_{210} + (V_{21j})$$

# Summary: Clustered Longitudinal Models

- Estimating 3-level models requires no new concepts, but everything is just at an order of complexity higher:
  - Proportioning variance over 3 levels instead of 2 → 2+ ICCs
  - Random slope variance will come from term directly beneath:
    - Level-2 random slope comes from level-1 residual
    - Level-3 random slope comes from level-2 random slope (or residual)
  - Level-1 effects can be random over level 2, level 3, or both
    - ICCs can be computed for level-1 slopes that are random over both level-2 and level-3 (assuming the L2 and L3 models match)
    - Convergence of level-1 effects should be tested over levels 2 AND 3
  - Level-2 effects can be random over level 3
    - Convergence of level-2 effects should be tested over level 3
  - Level-3 effects cannot be random; no convergence testing needed
  - Phew….

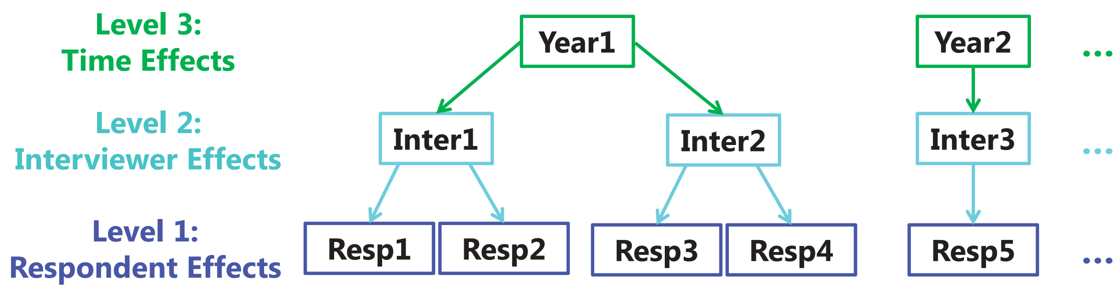


# Three-Level Models for Clustered Longitudinal Data

- Topics:
  - Decomposing variation across three levels in clustered longitudinal data
  - Unconditional (time only) model specification
  - Conditional (other predictors) model specification
  - **Other kinds of three-level designs**

# Other 3-Level Designs

- The sampling design for the outcome (not the predictors) dictates what your levels will be, **so time may not always be level 1**
- Example: Predicting answer compliance in respondents nested in interviewers, collected over several years (all different people)

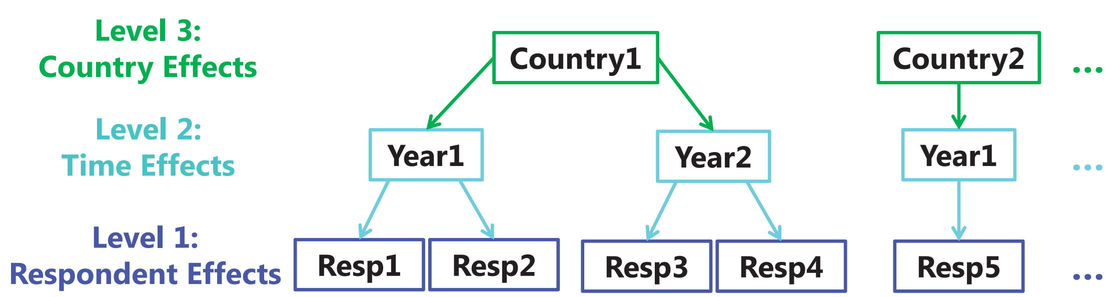



- Based on the sampling of time, time may be modeled…
  - As fixed effects in the model for the means → 2-level model instead
    - Best to use dummy codes for time if few occasions OR no time-level predictors of interest
  - As a random effect in the model for the variance → 3-level model
    - Then differences in compliance rates over time can be predicted by time-level predictors



# Other 3-Level Designs

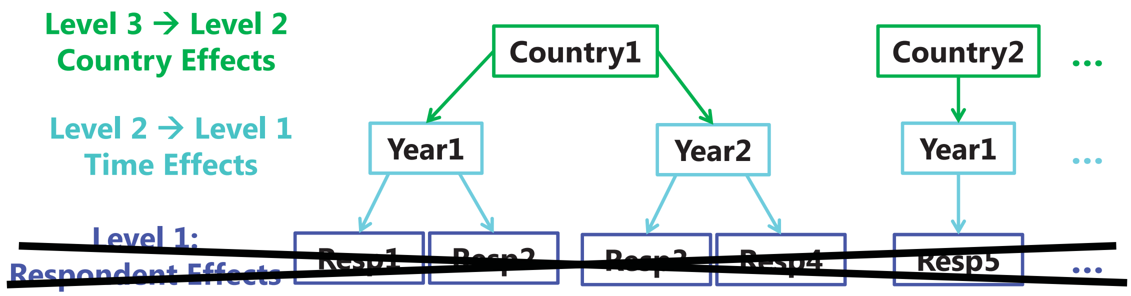
- Another example: Predicting **time-specific respondent outcomes** for people nested in countries, collected over several years (all different people, but the same countries measured over time)




- Before including any fixed effects of time, country and time are actually crossed, not nested as shown here
  - Are nested after controlling for which occasion is which via fixed effects (using dummy codes per mean or a time trend that describes the means)
  - Time is still a level because not all countries change the same way

# 3-Level Designs: Predictors vs. Outcomes

- Same example: What if, instead of respondent outcomes, we wanted to predict **time-varying country outcomes**?



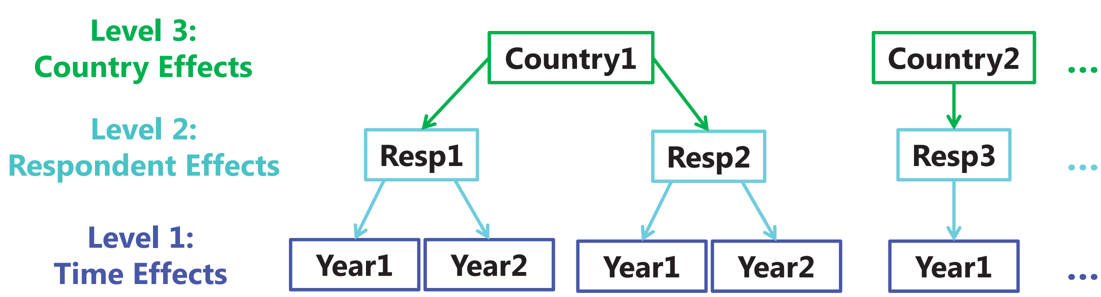

Because the outcome was measured at level 2 (country per time):

- Respondents are no longer a level at all (no outcomes for them)
- So there is nothing for respondent predictors to do, except at higher levels
  - **Time-specific averages** of respondent predictors → time-level outcome variation
  - **Across time, country averages** of respondent predictors → country-level outcome variation



# Other 3-Level Designs: Predictors by Level

- Last example: Predicting **time-specific respondent outcomes** for people nested in countries, collected over several years (all same people and same countries are measured over time)




- Country predictors can be included at level 3 only (no random effects)
- Person predictors should be included at levels 2 and 3 (+random over 3)
- What about effects of time-varying predictors?
  - For People: effects should be included at all 3 levels (+random over 2 and 3)
  - For Countries: effects are only possible at levels 1 and 3 (+random over 3)