Applied Multilevel Models

A Workshop Prepared for the
Institute for Social Research
Population Dynamics and Health Program
University of Michigan

Ryan W. Walters, PhD
November 30, 2022

Your Workshop Instructor

Ryan W. Walters, PhD
Associate Professor
Vice Chair for Clinical Research
Director of Statistics and Informatics
Department of Clinical Research and Public Health
Creighton University
RyanWalters@creighton.edu

Applied Multilevel Models

Part 1 of 12:
Introducing that Multilevel Model

What is a Multilevel Model (MLM)?

- Approximate synonyms
  - Generalized Linear Mixed Model (statistics)
    - "Mixed" implies fixed and random effects
  - Random Coefficient Model (statistics)
    - Random coefficients = random effects = latent variables
  - Hierarchical Linear Model (education)
    - Not the same as hierarchical regression

- Types of MLMs (clustered designs) ← This is where we live for today's workshop
  - Random-effects ANOVA
  - Clustered/nested observations model (e.g., kids in schools)
  - Cross-classified models (e.g., value-added models)
  - Psychometric models (e.g., factor analysis, IRT)

- Types of MLMs (longitudinal designs)
  - Repeated-measures ANOVA
  - (Latent) Growth curve model (Latent implies SEM)
  - Within-person fluctuation model (e.g., daily diary study)

Two Sides of Any Statistical Model

- Model for the Means (Structural)
  - Fixed effects
  - What you are used to caring about for hypothesis tests
  - How expected outcome for a given observation varies as a function of predictor variables

- Model for the Variance (Stochastic)
  - Random effects and residuals
  - What you are used to making assumptions about
  - How errors are distributed across observations (e.g., person, groups, etc.)
    - These relationships are generally called "dependency" and are the primary way MLM differs from regression and ANOVA

Dimensions for Organizing Models

- Outcome type: (conditionally) normal vs. not normal
- Dimensions of sampling: One vs. multiple

- Generalized Linear Models*
  - Any conditional outcome distribution
  - Fixed effects only through link functions
  - One dimension of sampling

- Generalized Linear Mixed Models (GLMM)
  - Any conditional outcome distribution
  - Fixed and Random effects through link functions
  - Multiple dimensions of sampling

Know that GLMMs subsume the multilevel model

*__Note 1__: "General Linear Model" = identity link, normal distribution
*__Note 2__: Least squares can only be used for General Linear Model

What can MLM do for You?

1. Model dependency across observations
   - Clustered or cross-classified data? No problem!

2. Include predictors on any scale at any level
   - Person-level or group-level predictors
   - Explore reasons for dependency, don't just control for it

3. Does not require same data structure for each group
   - Unbalanced or missing data? No problem! (with caveat)

4. You already know how (you'll know more soon)!
   - Use Stata, SAS, SPSS, R, Mplus, HLM, MLwiN
   - What's an intercept?
   - What's a slope?
   - What's variance component?

1. Model Dependency

- Source(s) of dependency depend on sources of variation created by your sampling design
  - Residuals for outcomes from the same clustering unit are likely to be related, which violates assumption of independence

- Levels of dependency = levels of random effects
  - Sampling dimensions can be **nested** (e.g., people within groups)
  - No clean nested structure? Likely a **crossed** sampling design
    - e.g., kids in neighborhoods who attend different schools
  - To have a level, there must be random outcome variation due to sampling that **remains** after including fixed effects
    - But, could have fixed effects explain all random variation (the goal, right?)

Dependency is Created by…

- Mean differences across higher-level sampling units (i.e., groups)
  - Constant between-group dependency/correlation
  - Quantified by the random intercept
- Between-group differences in the effect of person-level predictors
  - Non-constant dependency in the size of the fixed effect across groups
  - Represented by random slopes
- Non-constant within-group correlation for unknown reasons
  - Autocorrelation (e.g., AR1 structure)
  - Generally, this does not apply to clustered data as people at level 1 are assumed to be exchangeable

Should We Care About Dependency?

- Say we have a wrong Model for the Variance
  - i.e., wrongly assume independence
- Validity of the tests of predictors depend on having the "right" (err, least wrong) model for the variance
  - Estimates will usually be okay, but standard errors for these estimates (and, thus, *p*-values) will likely be biased
- The sources of variation in your outcome will dictate what kind of predictors are most useful
  - Between-group variation at level 2 require group-level predictors
  - Within-group variation at level 1 require within-group predictors

2. Include Predictors at Any Level of Analysis

- ANOVA
  - Test differences among discrete IV factors/levels/conditions
  - Continuous covariates, too? Get some ANCOVA or just use…
- Regression
  - Test whether slopes relating predictors to outcomes differ from 0
- What if predictor values differ across people at level 1 but can't be characterized by conditions (e.g., continuous age)?
  - We have to use the multilevel modeling framework

Let's Talk about Predictors

- Variables that are constant (invariant) within a level-2 group
  - i.e., values are not different within a group <u>during the study period</u>
  - e.g., right-to-work law status for a given calendar year

- Variables that are not constant within a level-2 group
  - Variables that have different values across level-1 persons
  - e.g., age, biological sex

- Some predictors might only be measured at higher levels
  - e.g., person unemployed vs. state unemployment

- Interactions between levels can be included, too
  - Does the effect of <u>age</u> differ by mid-year <u>unemployment rate</u>?

Level: Person Group

3. Does not require same data structure per group

- Multilevel models use stacked (aka, long) data structure
  - Rows missing data are excluded
- Consider the data on the right
  - DV = Disability
  - IVs = Male, Unemployment
- If only using Male as a predictor, then State 2 excludes person 4
- If using only Unemployment as a predictor, then State 3 is excluded

StateID	PersonID	Male	Unemployment	Disability
1	1	0	2.2	1
1	2	1	2.2	4
1	3	0	2.2	3
2	1	1	5.6	0
2	2	0	5.6	5
2	3	0	5.6	2
2	4	.	5.6	6
3	1	0	.	3
3	2	0	.	4
3	3	1	.	2

4. You already know how!

- If you can do ANOVA/regression, you can do multilevel models. Period. Trust me.

$$Weight_i = 150 - 2(Age_i) + 50(Male_i) + e_i$$

- How do you interpret the estimate for…
  - The intercept?
  - The effect (slope) of continuous predictor age?
  - The effect (slope) of categorical predictor sex?
  - The residual value?
  - The residual variance (e.g., variance component)

Applied Multilevel Models

Part 2 of 12:

Statistical Approaches for Clustered Data

Brief Comment: Fixed Effects Models

- You can use fixed or random effects to "handle" between-group correlation

- The fixed effects approach explains group-specific differences
  - Include n_{group} – 1 binary group-indicator variables as fixed effects
  - This approach uses those fixed effects to control for sampling/dependency
  - Allows inferences about group differences (the end)

- Problem abound
  - No additional group-level predictors can be included
  - Those indicator variables explained all the reason why groups differ
  - There is no remaining between-group variance

- Recommended approach if you have <10 groups (meh…)
  - Has to do with precision of estimated random effect variances from a multilevel model

Example Time!

Example - Fixed Effects.pdf

(Introduction of Working Example and Models 1-3)

Enter the Multilevel Model

- Builds on the ANOVA/regression framework

- Quantify between-group differences via random effects
  - Directly measure how much of the outcome variance is due to between-group differences

- Predict between-group differences via fixed effects
  - Include fixed effects of predictors at any level of analysis

- Random effects give you predictable control of dependency

Data Requirements of the Multilevel Model

- Multiple **outcomes** from same sampling unit
  - More data is better (with diminishing returns)

- Any measurement scale can be analyzed using appropriate link functions and (conditional) response distributions
  - We will focus on interval scale (conditionally normal)
  - Scores must hold the same meaning across all observations
    - Implies measurement invariance
    - Includes the meaning of the construct

- Oh, and fancy statistical models **cannot** save badly measured variables or junky research designs!

Levels of Inference for Multilevel Data

- Between-group (BG) Relationships
  - Level-2 = group-level = "INTER-group Differences"

- Within-Group (WG) Relationships
  - Level-1 = Person-level = "INTRA-group Variation"

- Multilevel models allow examination of both types of relationships simultaneously
  - This is important (see Figure)
  - Be aware that most person-level predictors usually have level-1 and level-2sources of variation!

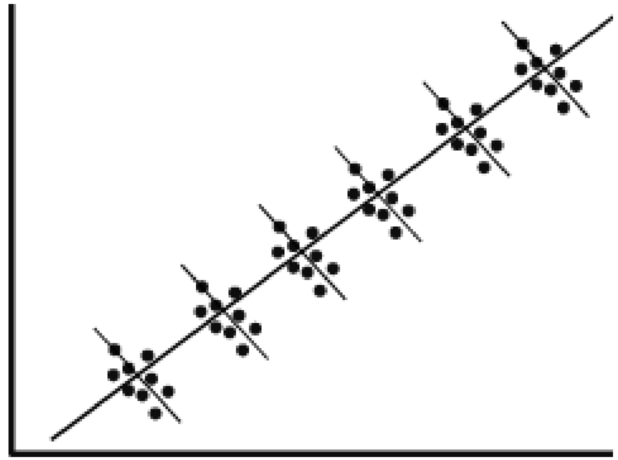


Figure stolen from: https://stats.oarc.ucla.edu/other/mult-pkg/introduction-to-linear-mixed-models/

Applied Multilevel Models

Part 3 of 12:
Visualizing the Multilevel Model

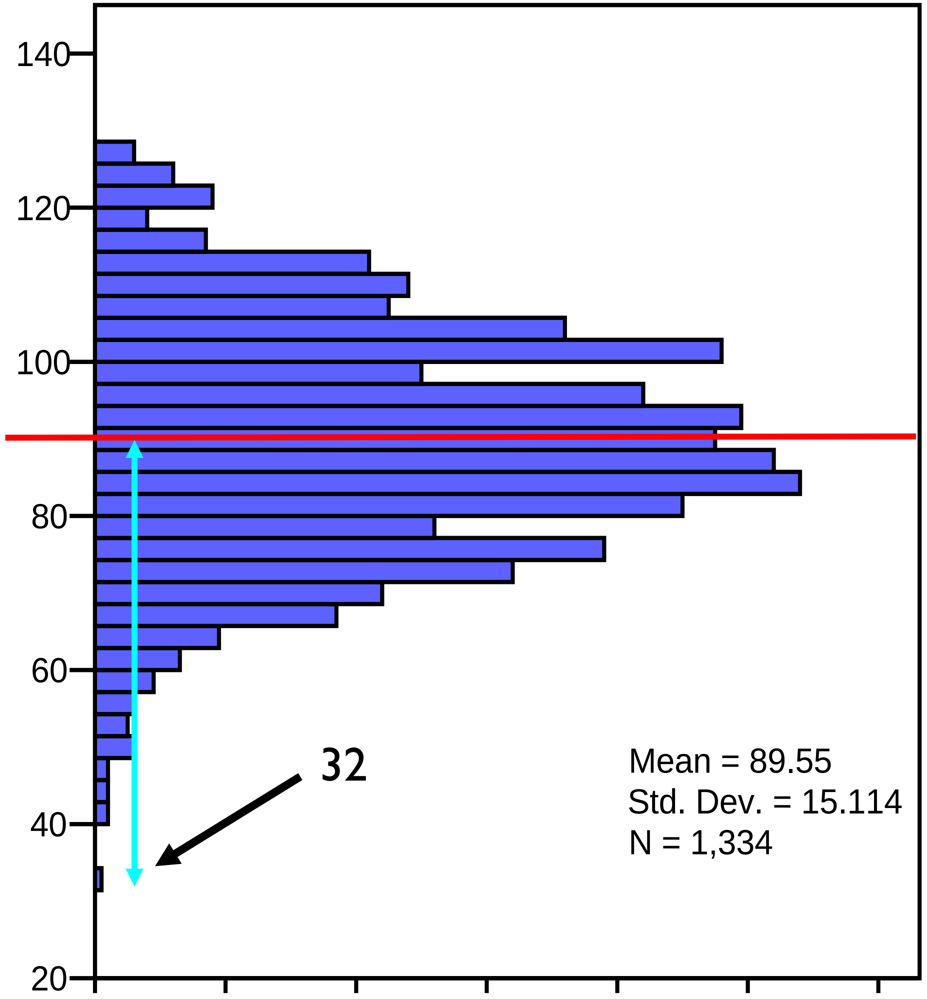
A Little Between-group Model

$$y_j = \boxed{\beta_0 + \beta_1 X_j + \beta_2 Z_j + \beta_3 X_j Z_j} + \boxed{e_j}$$

- Model for the Means
  - The outcome y_j is measured once per group (subscript j = group)
  - Each group's expected (predicted) outcome weighted by a linear combination of their values on X_j, Z_j, and their interaction $X_j Z_j$
  - Estimated parameters $\beta_0, \beta_1, \beta_2, \beta_3$ are called fixed effects because they apply equally to every group in the sample

- Model for the Variance
  - One residual e_j for each group
  - The e_j across groups are assumed independent and normally distributed with a mean of 0 with constant estimated variance σ_e^2…that is: $e_j \sim N(0, \sigma_e^2)$
  - Note that the residual variance σ_e^2 is the only estimated variance component in a between-group model

The Unconditional Between-Group Model



$$y_j = \beta_0 + e_j$$

For one group j:

$$32 = \underbrace{89.55}_{\hat{y}_j} + (-58)$$

Model for the Means

Error Variance of y

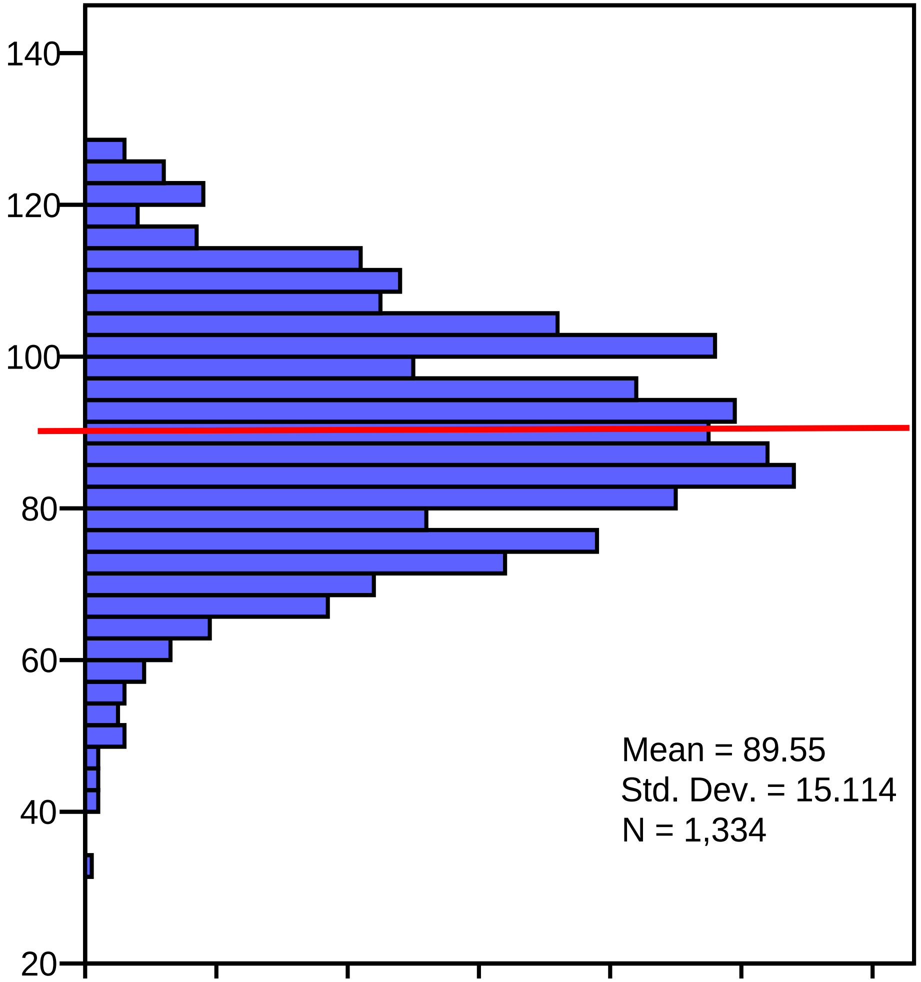
$$\frac{\sum_{i=1}^{N}(y_j - \hat{y}_j)^2}{N-1}$$

Model for the Variance

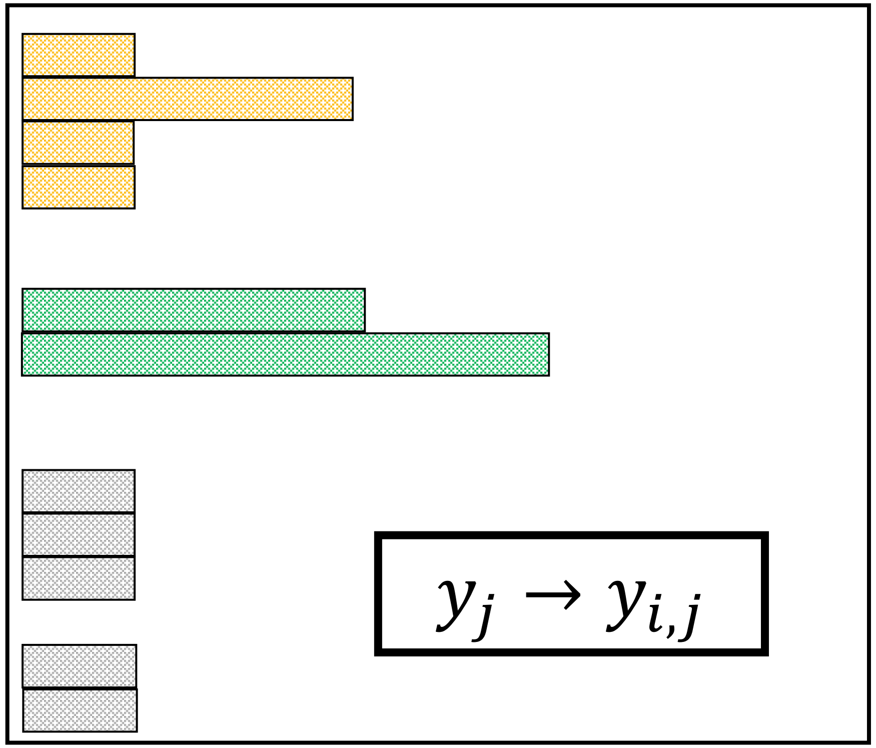
$$\sigma_e^2 = (15.114)^2 = 228.43$$

Let's Sprinkle in Some Within-group Information

Full Sample Distribution

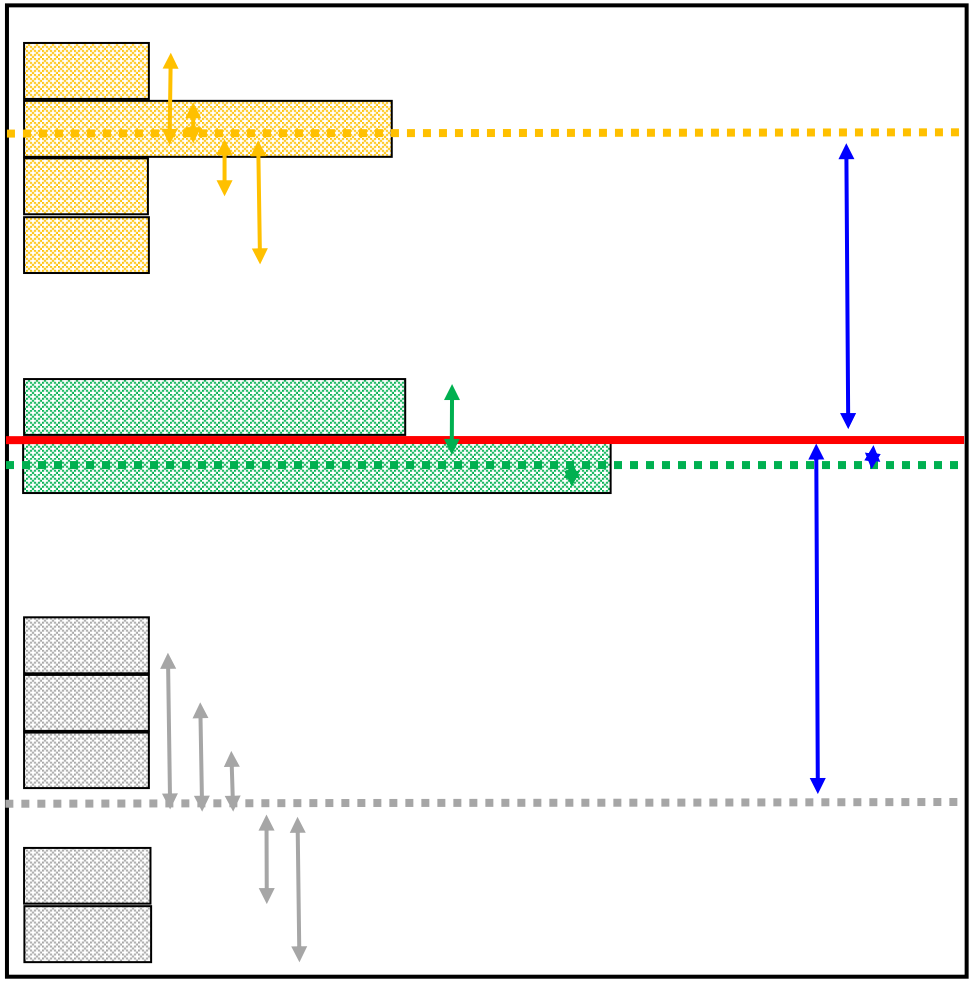


3 Groups of 5 People



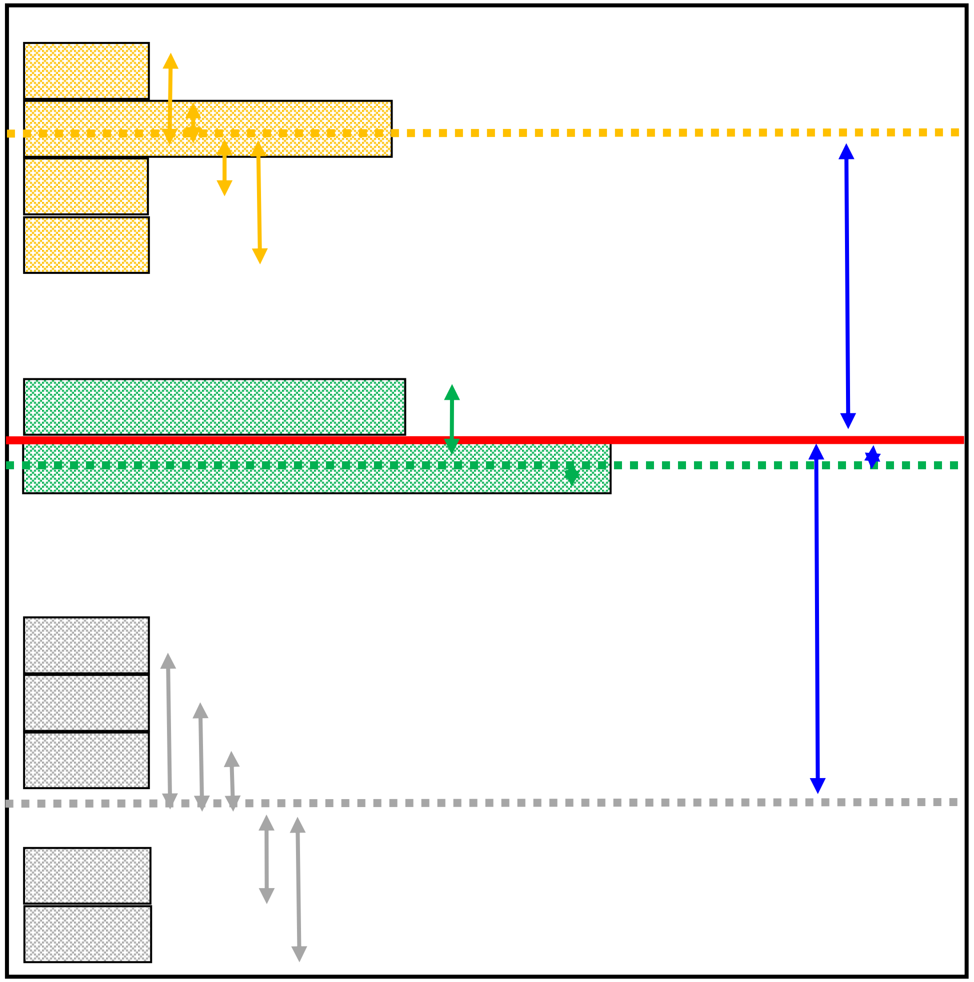
$$y_j \rightarrow y_{i,j}$$

Unconditional Between- & Within-group Model



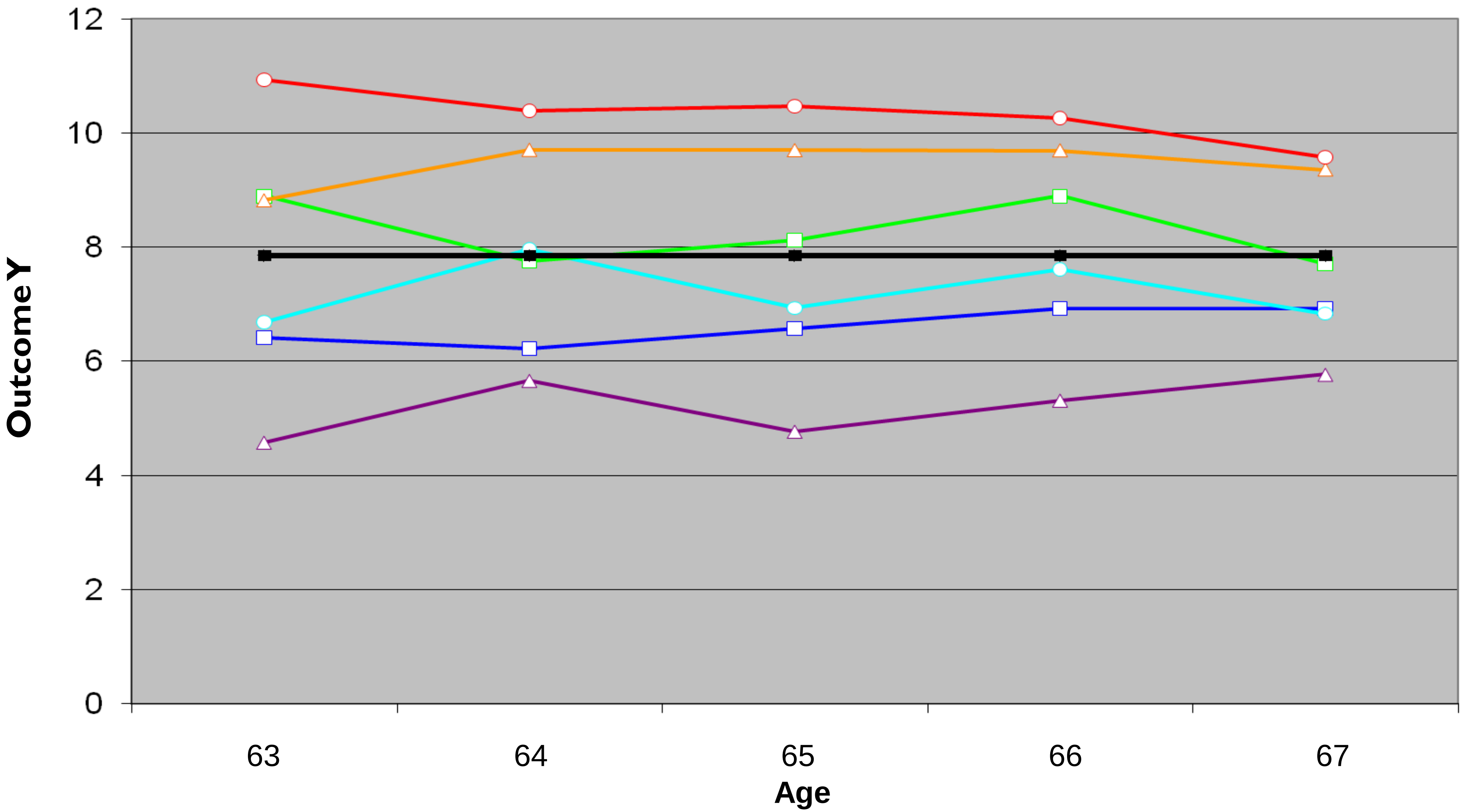
- Start with mean of $y_{i,j}$ as "best guess"
  - = the grand mean across all observations
  - = the fixed intercept, β_0
- Better guess by considering persons within a group
  - = group mean, $\beta_{0,j}$
- Deviations: $\beta_{0,j} - \beta_0$
  - = random intercept, $U_{0,j}$
- Deviations: $y_{i,j} - \beta_{0,j}$
  - = residual, $e_{i,j}$

Unconditional Between- & Within-group Model

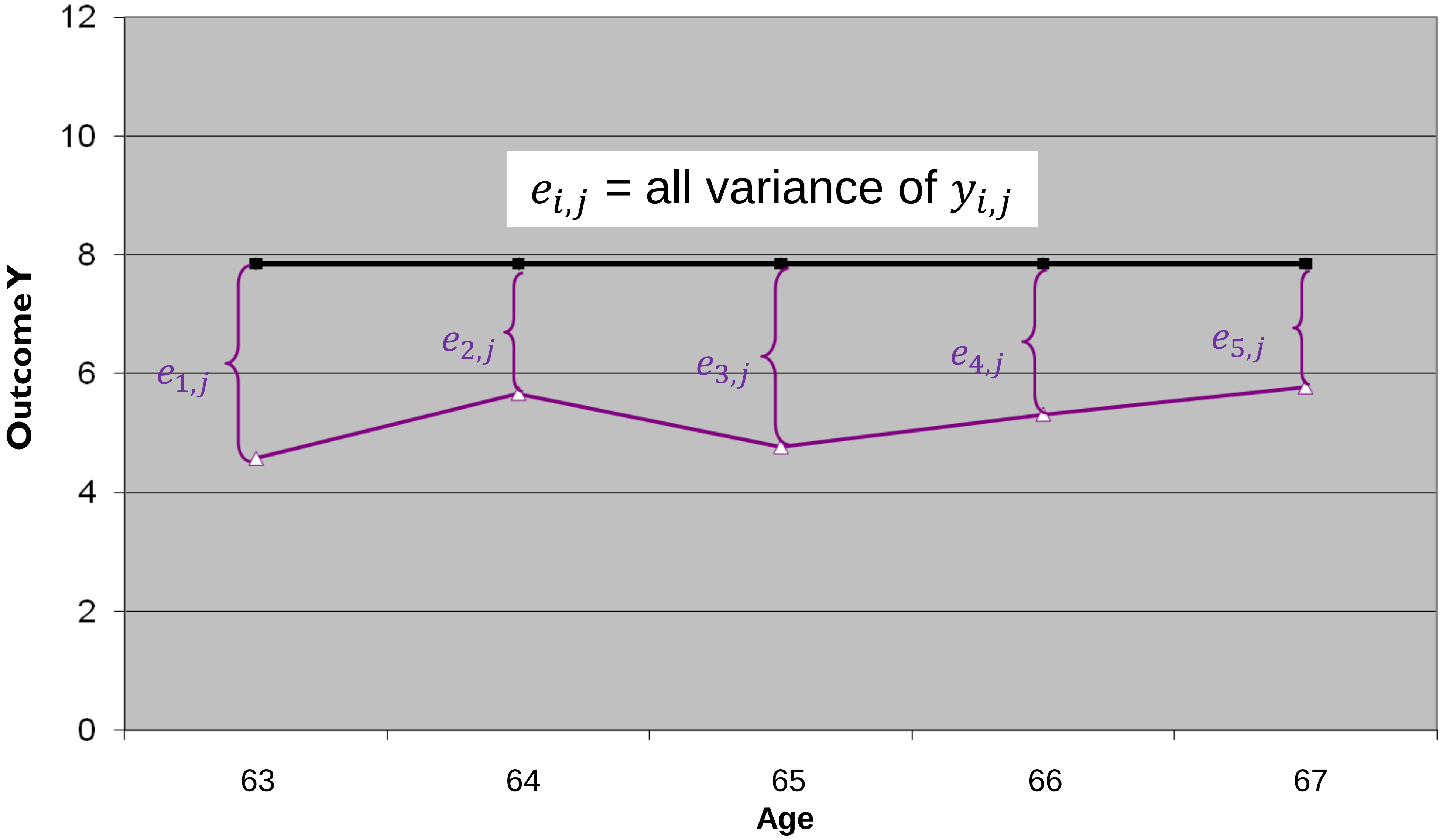


- Total variance of $y_{i,j}$ now has two sources
  - i.e., two variance components
- Between-group variance ($\tau_{U_0}^2$)
  - Deviations of group-specific mean from the fixed intercept
  - Random intercepts: $U_{0,j}$
- Within-group variance (σ_e^2)
  - Deviations of a person's observation from their groups' mean
  - Residuals: $e_{i,j}$

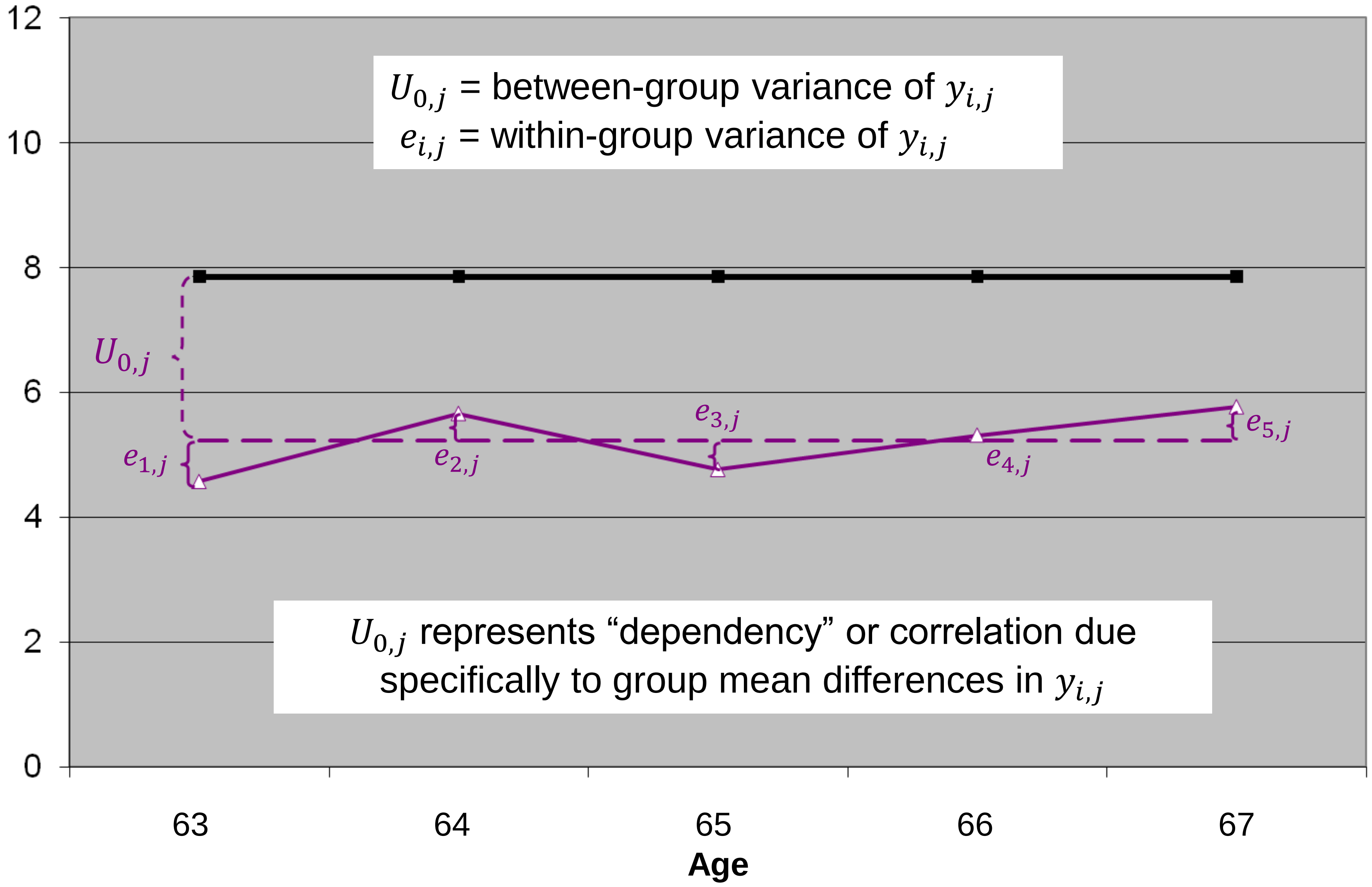
Let's Consider Data from Six Groups



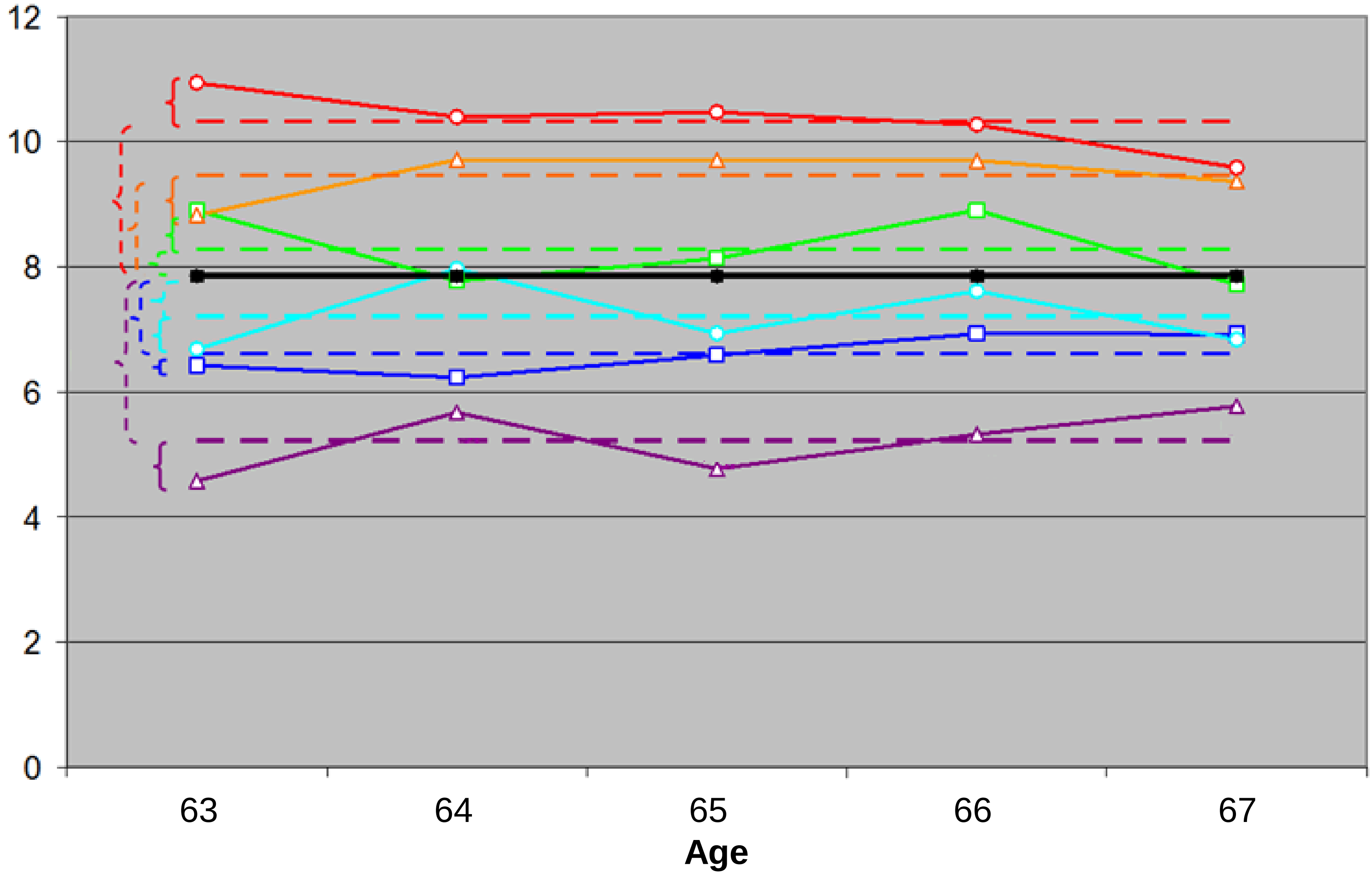
Unconditional Between-group Model for Group j



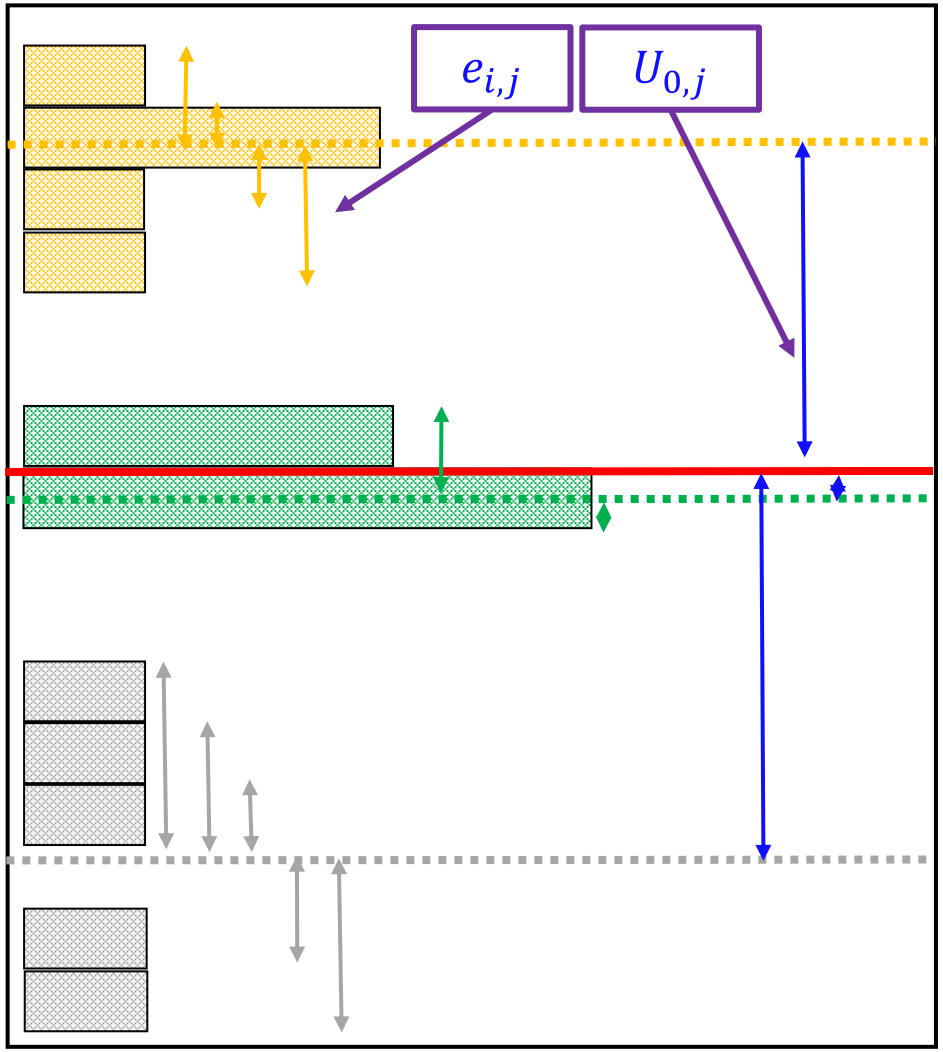
Unconditional Between- & Within-group Model for Group j



Unconditional Between- & Within-group Model for Six Groups



Unconditional Between- & Within-Person Model



- Total variance of $y_{i,j}$ has two variance components

- Level-2 random intercept variance
  - Variance of the $U_{0,j}$ as $\tau_{U_0}^2$
  - Between-group variance
  - $U_{0,j}$ = group-specific deviations from the **fixed** intercept

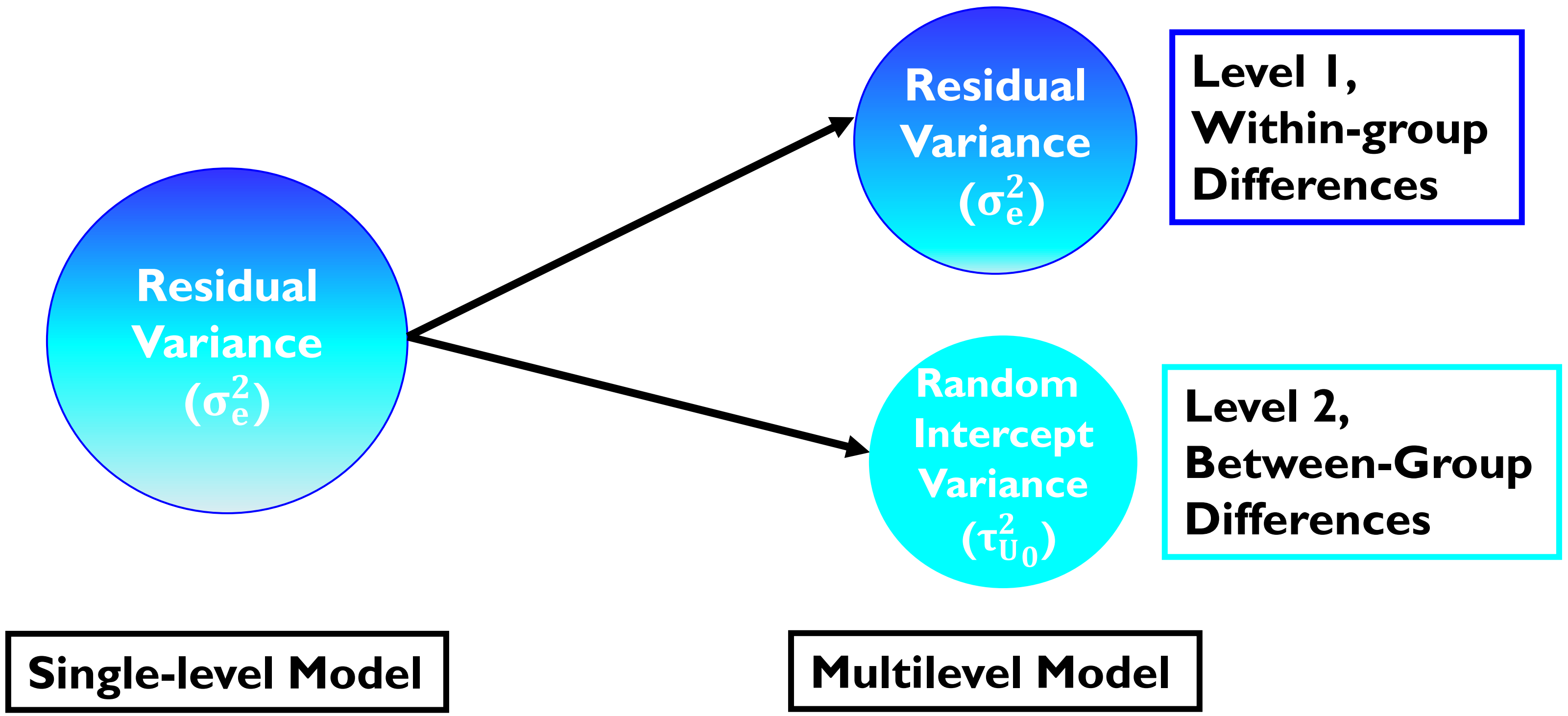
- Level-1 residual variance
  - Variance of the $e_{i,j}$ as σ_e^2
  - Within-group variance
  - $e_{i,j}$ = deviations from person's group mean

Applied Multilevel Models

Part 4 of 12:
The Intraclass Correlation

How the Multilevel Model Handles Dependency

- The multilevel model "handles" correlated data
  - But where does that correlation go?
  - Into a new variance component that is partitioned out of residual variance
  - Note: partitioning variance ≠ explaining variance!!
  - Only predictor variables (i.e., fixed effects) explain variance



Empty Means, Random Intercept Model

- Model for the Means
  - 1 parameter
    - Fixed intercept $\gamma_{0,0}$
- Model for the Variance
  - 2 parameters
    - Level-1 residual variance σ_e^2
    - Level-2 random intercept variance $\tau_{U_0}^2$

- Empty single-level model

  $y_i = \beta_0 + e_i$

- Empty multilevel model
  - Level 1

    $y_{i,j} = \beta_{0,j} + e_{i,j}$

  - Level 2

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

  - Composite

    $y_{i,j} = (\gamma_{0,0} + U_{0,j}) + e_{i,j}$

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

Random intercept = group-specific deviation from predicted intercept

Fixed intercept = mean of group means given no predictors (yet)

Unconditional Intra-Class Correlation (ICC)

$$\text{ICC} = \frac{\text{Between−group Variance}}{\text{Between−group Variance} + \text{Within−group Variance}}$$

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

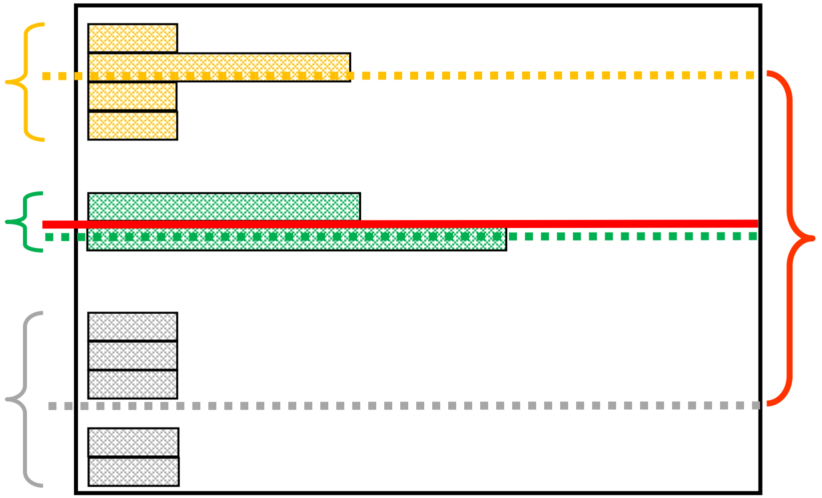
$$\text{ICC} = \frac{\tau_{U_0}^2}{\tau_{U_0}^2 + \sigma_e^2}$$

- The many definitions of ICC
  - Proportion of total variance that is between groups
  - Average correlation among persons (not a fan of this as ICC > 0)
  - Effect size for constant group dependency

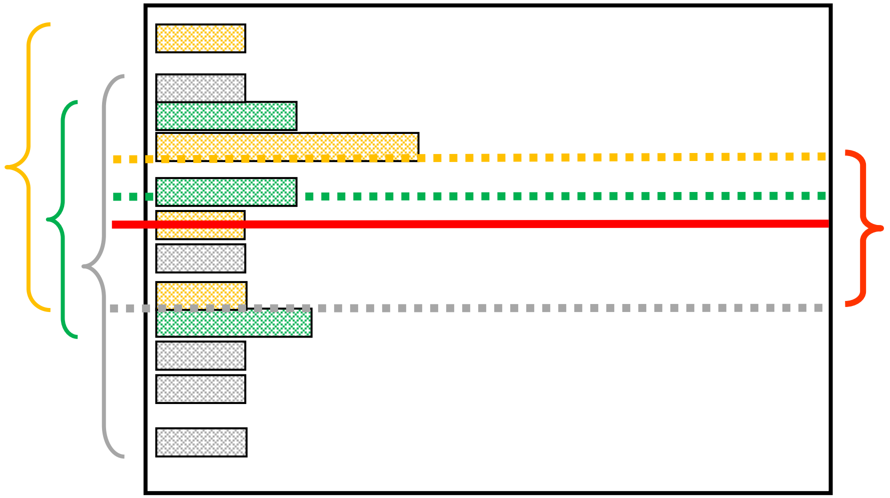
- The ICC quantifies how badly we need to worry about dependency
  - As in, an approximation of how wrong could be we if dependency was ignored

Unconditional Intra-Class Correlation (ICC)

$$\text{ICC} = \frac{\text{Between-group Variance}}{\text{Between-group Variance} + \text{Within-group Variance}} = \frac{\tau_{U_0}^2}{\tau_{U_0}^2 + \sigma_e^2}$$



- Large(r) ICC
  - Random intercept variance larger than residual variance



- Small(er) ICC
  - Residual variance larger than random intercept variance

RyanWalters@creighton.edu

Can We Ignore Clustering if ICC ~ 0?

- There is no value of ICC that is uniformly "safe" to ignore because…

- Unconditional and conditional ICCs will differ
  - "Conditional" indicates after predictors are included
  - The purpose of predictors is to explain (i.e., reduce) variance and explaining variance = changing ICC

- Too, reducing the residual variance often results in an increase in the random intercept variance, which then increases the conditional ICC

$$\text{Estimated } \tau^2_{U_0} = \text{True } \tau^2_{U_0} + \left(\frac{\sigma^2_e}{n_{L1}}\right)$$

$$\text{True } \tau^2_{U_0} = \text{Estimated } \tau^2_{U_0} - \left(\frac{\sigma^2_e}{n_{L1}}\right)$$

- Takeaway: just use a multilevel model

Need an MLM? Use Model Comparison

- Testing ICC > 0 requires model comparison
- Relative model fit is indexed by a "deviance" statistic = –2LL
  - Labeled as -2 log likelihood in SAS and SPSS, but given as LL in Stata and R
  - Measure badness of fit, so smaller values are better
- Two estimation flavors: Maximum Likelihood (ML) or Residual ML (REML)
  - If using REML, the predictor variables must be identical between comparisons models
- Significance determined by –2ΔLL test (aka, likelihood ratio test, deviance difference test)
  1. Calculate –2ΔLL: $(-2LL_{\text{fewer}}) - (-2LL_{\text{more}})$
  2. Calculate Δdf: $(\#\ \text{parameters}_{\text{more}}) - (\#\ \text{of parameters}_{\text{fewer}})$
  3. Compare –2ΔLL to critical values of χ^2 distribution with df = Δdf
- Musings
  - –2LL is summed across observations; make sure sample size is equal between comparison models
  - Add parameters: model is better or not better; remove parameters: model is worse or not worse

Quantifying Random Effect Variances

- The ICC quantifies the proportion of variance at level 2, but…
- If a random effect variance is statistically significant, its interpretation is meaningless without context
  - Consider a fixed intercept ($\gamma_{0,0}$) = 1.966 with a significant random intercept variance ($\tau^2_{U_0}$) = 0.005
  - What does that 0.005 indicate in terms of the actual outcome?
- Enter the 95% random effects confidence interval (RECI)
  - Can calculate for every random effect variance in your model
  - Provides a range of values around the fixed effect that captures 95% of level-2 groups

$$95\%\ \text{RECI} = \text{fixed effect} \pm (1.96 * \sqrt{\text{random effect variance}})$$

$$95\%\ \text{RECI for random intercept} = 1.966 \pm (1.96 * \sqrt{0.005}) = [1.827, 2.105]$$

- So, groups are predicted to have an outcome of 1.966 on average, but 95% of group-specific intercepts ranged from 1.827 to 2.105

Applied Multilevel Models

Part 5 of 12:

Model Assumptions via Matrices

Conditional MLM via Matrices

$$y_{i,j} = (\beta_0 + \beta_1 X_j + \beta_2 Z_j + \cdots) + U_{0,j} + e_{i,j}$$

$$\mathbf{Y}_j = \boxed{\mathbf{X}_j \boldsymbol{\beta}} + \boxed{\mathbf{Z}_j \mathbf{U}_j + \mathbf{e}_j}$$

$\boldsymbol{\beta}$ will eventually become $\boldsymbol{\gamma}$

- $\mathbf{Y}_j$ = n_i x 1 vector of outcomes for group j
- $\mathbf{X}_j$ = n_i x p design matrix of predictor variables for group j
- $\boldsymbol{\beta}$ = p x 1 vector of fixed effects (no subscript!)
- $\mathbf{Z}_j$ = n_i x q design matrix of predictor variables with random effects for group j
- $\mathbf{U}_j$ = q x 1 vector of random effects for group j
- $\mathbf{e}_j$ = n_i x n_i matrix of residuals for group j

- Both $\mathbf{X}_j$ and $\mathbf{Z}_j$ have first column of 1s to represent the intercept and it becomes a multilevel model when $\mathbf{Z}_j$ is subsumed within $\mathbf{X}_j$

Model Assumptions via Matrices

- Conditionally normally distributed outcome

$$\mathbf{Y}_j|\mathbf{U}_j \sim \mathbf{N}_{n_j}\left(\boldsymbol{\mu}_j, \mathbf{R} = \sigma_e^2 \mathbf{I}_{n_j}\right)$$

  - $\mathbf{I}_{n_j}$ = diagonal matrix of 1s with 0s on off-diagonal

- Residual assumptions (usually independent)

$$\mathbf{e}_j \sim \mathbf{N}_{n_j}(\mathbf{0}, \mathbf{R})$$

- Random effect assumptions (usually unstructured)

$$\mathbf{U}_j \sim \mathbf{N}_q(\mathbf{0}, \mathbf{G})$$

Typical **R** and **G** for Clustered Data

- For <u>one group</u>, say we have <u>four</u> persons and a random intercept

$$\mathbf{R} = \sigma_e^2 \mathbf{I}_{n_j} = \begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & \sigma_e^2 \end{bmatrix} \qquad \mathbf{G} = \left[\tau_{U_0}^2\right]$$

$$\mathbf{V} = \mathbf{ZGZ}^{\mathrm{T}} + \mathbf{R} = \begin{bmatrix} \sigma_e^2 + \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \sigma_e^2 + \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \tau_{U_0}^2 & \sigma_e^2 + \tau_{U_0}^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 & \sigma_e^2 + \tau_{U_0}^2 \end{bmatrix}$$

- The actual analysis uses the **V**-matrix
  - With just a random intercept, **V** is termed compound symmetric
  - The **V** <u>correlation</u> matrix provides ICC on the off-diagonal

Example Time!

Example - Grand-mean-centering.pdf

Example - Grand-mean-centering.xlsx

(Model 1)

Applied Multilevel Models

Part 6 of 12:

Adding Level-2, Group-level Predictors

Group-level Predictors

- When included by themselves (i.e., no interaction), they serve to moderate the intercept
- They are constant value within a group
- Important:
  - If their value is missing, the entire group is omitted from analysis (i.e., listwise deletion)
  - See State 3

StateID	PersonID	Male	Unemployment	Disability
1	1	0	2.2	1
1	2	1	2.2	4
1	3	0	2.2	3
2	1	1	5.6	0
2	2	0	5.6	5
2	3	0	5.6	2
2	4	.	5.6	6
3	1	0	.	3
3	2	0	.	4
3	3	1	.	2

Unemployment as a Level-2 Predictor

- Consider a state's mid-year unemployment rate that we center at 2%
  - $\text{SMue2}_j = \text{Unemployment}_j - 2$
  - i.e., 0 = 2%

- Level-1
  $y_{i,j} = \beta_{0,j} + e_{i,j}$

- Level-2 (one per β)
  $\beta_{0,j} = \gamma_{0,0} + \gamma_{0,1}(\text{SMue2}_j) + U_{0,j}$

- Composite
  $y_{i,j} = \gamma_{0,0} + \gamma_{0,1}(\text{SMue2}_j) + U_{0,j} + e_{i,j}$

$\gamma_{0,0}$ = the predicted outcome for a state with 2% unemployment

$\gamma_{0,1}$ = the difference in the average outcome for groups that average one-percent higher unemployment

Variance Explained by Level-2 Predictors

- Quantify variance explained using pseudo-R^2
  - A pseudo-R^2 can be calculated for every variance component
  - Problem! Variance components shift around so pseudo-R^2 can be negative
  - Negative pseudo-R^2 is more common with REML
  - Hard to explain to readers, but if pseudo-R^2 is negative, just call it 0

- Fixed effects of level-2 predictors by themselves
  - Level-2 main effects and interaction effects reduce level-2 random intercept variance

$$\text{Pseudo}-R^2_{\tau^2_{U_0}} = \frac{\text{random intercept variance}_1 - \text{random intercept variance}_2}{\text{random intercept variance}_1}$$

- An alternative is Total R^2
  - Quantifies the <u>total</u> variance explained across levels
  - Akin to R^2 from garden-variety linear regression

1. Get model-predicted outcome from fixed effects (<u>not</u> including random effects)
2. Get the Pearson correlation between model-predicted and observed outcome
3. Square that correlation and you've got yourselves a total R^2

Example Time!

Example - Grand-mean-centering.pdf

Example - Grand-mean-centering.xlsx

(Models 2a-2b)

Applied Multilevel Models

Part 7 of 12:

(the complexity of) Adding Level-1, Person-level Predictors

The Joy of Level-1 Predictors

- Modeling level-1 predictors is complicated. Period.

- They represent an aggregated effect of two sources of variance
  - Between-group (BG): some groups average more of the predictor than other groups
  - Within-group (WG): some people have more predictor than others in their group

- There is no conceptual difference between the outcome and a level-1 predictor
  - Remember, your outcome might be someone else's covariate

- We quantify BG and WG variation for the level-1 predictor using the ICC
  - ICC = between / (between + within)
  - ICC > 0: the level-1 predictor has between-group variation
  - ICC < 1: the level-1 predictor has within-group variation

Between-group vs. Within-group Effects

- Consider student and school SES on achievement…
  - BG: Schools with more rich kids may have greater mean achievement than schools with more poor kids
  - WG: Rich students in a school may have greater achievement than poor students in that school

- Variable partitions can have different scales at different levels
  - Level-1: student biological sex (0 = male; 1 = female)
  - Level-2: school percent of female students (range: 0% to 100%)

- There are two centering options to disaggregate the level-1 and level-2 effects of the level-1 predictor
  - Grand-mean-centering
  - Group-mean-centering

- Level-1 centering choice dictates level-2 interpretation

Modeling Level-1 Predictors

- Consider people clustered in states and level-1 predictor $\text{Age}_{i,j}$

- Level-2, between-group effect of $\text{Age}_{i,j}$
  - Represented by the state-specific mean of $\text{Age}_{t,i}$ $(\text{SMage}_j = \overline{\text{Age}}_j)$
  - Is the average age of the state (based on the sampled data)
  - Center SMage_j to ensure a meaningful 0 $(\text{SMage65}_j = \text{SMage}_j - 65)$

- Level-1, within-group effect is based on centering choice…
  - Grand-mean-centering
    - Center $\text{Age}_{i,j}$ at some constant value (e.g., $\text{Age65}_{i,j} = \text{Age}_{i,j} - 65$)
    - Here, $\text{Age65}_{i,j}$ still retains level-1 and level-2 variability
  - Group-mean-centering
    - Center $\text{Age}_{i,j}$ at their state's mean age $(\text{WSage}_{i,j} = \text{Age}_{i,j} - \text{SMage}_j)$
    - Here, $\text{WSage}_{i,j}$ has a pure level-1 effect
    - Literally, subtract off the level-2 age effect

- The interpretation of the level-1 and level-2 fixed age effects differ based on the centering choice

Level-1 Predictors Contain Three Effects

- Level-1, within-group effect
  - If you have higher predictor values than others in your group, do you also have higher outcome values than others in your group?
  - Effect explains level-1, residual variance (σ_e^2)

- Level-2, between-group effect
  - Do groups who average higher predictor values compared to other groups also average higher outcome values?
  - Effect explains level-2, random intercept variance ($\tau_{U_0}^2$)

- Level-2, contextual effect
  - After controlling for the value of the level-1 predictor for each person, is there an incremental contribution of averaging higher predictor values?
  - Do the level-2 between-group and level-1 within-group effects differ?
  - If no contextual effect, then level-2 = level-1 (termed *convergence*)
  - Effect explains level-2, random intercept variance ($\tau_{U_0}^2$)

- Either centering decision will only provide two of the three effects…

Applied Multilevel Models

Part 8 of 12:

Grand-mean-centering

Why Not Include a Level-1 Predictor by Itself?

- Consider $\text{Age65}_{i,j}$ included in the model by itself

- Level 1

  $y_{i,j} = \beta_{0,j} + \beta_{1,j}(\text{Age65}_{i,j}) + e_{i,j}$

$\text{Age65}_{i,j} = \text{Age}_{i,j} - 65$
Has both within-group and between-group variation

- Level 2

  $\beta_{0,j} = \gamma_{0,0} + U_{0,j}$
  $\beta_{1,j} = \gamma_{1,0}$

Because $\text{Age65}_{i,j}$ still contains both BG and WG variation, its one fixed effect has to do the work of two predictors.
In a word: **Inaccurate**!

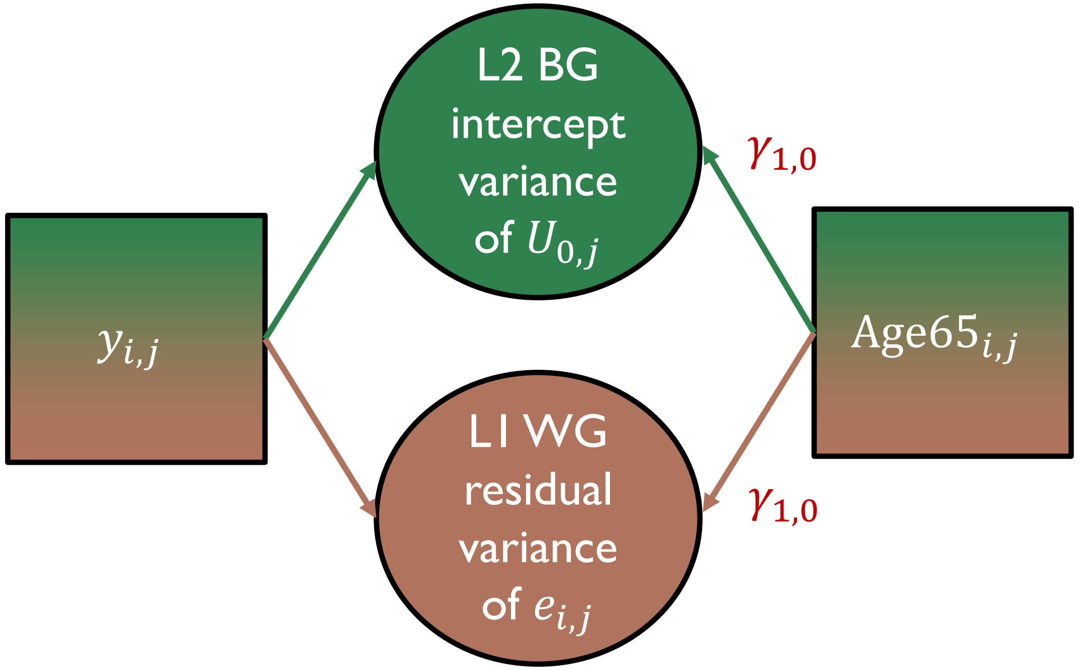
$\gamma_{1,0}$ = combined BG and WG effect!

If the level-1 predictor is included by itself, its fixed effect assumes convergence (i.e., level-1 = level-2).
This is testable via the contextual effect.
Synonyms for the combined effect include smushed, convergence, conflated, or composite effect.

Level-1 Predictor by Itself

Model-based partitioning of $y_{i,j}$ outcome variance

No partitioning of $\text{Age}_{i,j}$ so it only has one fixed effect that represents the combined BG and WG effects



- This will occur whenever the level-1 predictor has a non-zero ICC!
- Know that the convergence effect ($\gamma_{1,0}$) will often be closer to the within-group effect simply because there is more data at level-1

Variance Explained by Level-1 Predictors

- Fixed effects of level-1 predictors by themselves
  - Level-1 main effects explain level-1 residual variance (σ_e^2)
  - Level-1 interactions explain level-1 residual variance (σ_e^2)

$$\text{Pseudo}-R^2_{\sigma_e^2} = \frac{\text{residual variance}_1 - \text{residual variance}_2}{\text{residual variance}_1}$$

- When the level-1 effect retains both level-1 and level-2 variability, it will explain level-1 residual variance (σ_e^2) and level-2 random intercept variance ($\tau_{U_0}^2$)

- A single predictor that reduces variance across levels is a telltale sign the predictor effects need to be disaggregated

Disaggregating Level-1 from Level-2

- Within- and between-group effects are estimated <u>only</u> when both levels are included in the model
  - Include both level-1 $\text{Age65}_{i,j}$ and level-2 SMAge65_j

- Level 1
  $$y_{i,j} = \beta_{0,j} + \beta_{1,j}(\text{Age65}_{i,j}) + e_{i,j}$$

- Level 2
  $$\beta_{0,j} = \gamma_{0,0} + \gamma_{0,1}(\text{SMage65}_j) + U_{0,j}$$
  $$\beta_{1,j} = \gamma_{1,0}$$

$\text{Age65}_{i,j} = \text{Age}_{i,j} - 65$
Contains both between-group and within-group variation

$\text{SMage65}_j = \overline{Age}_j - 65$
Only between-group variation

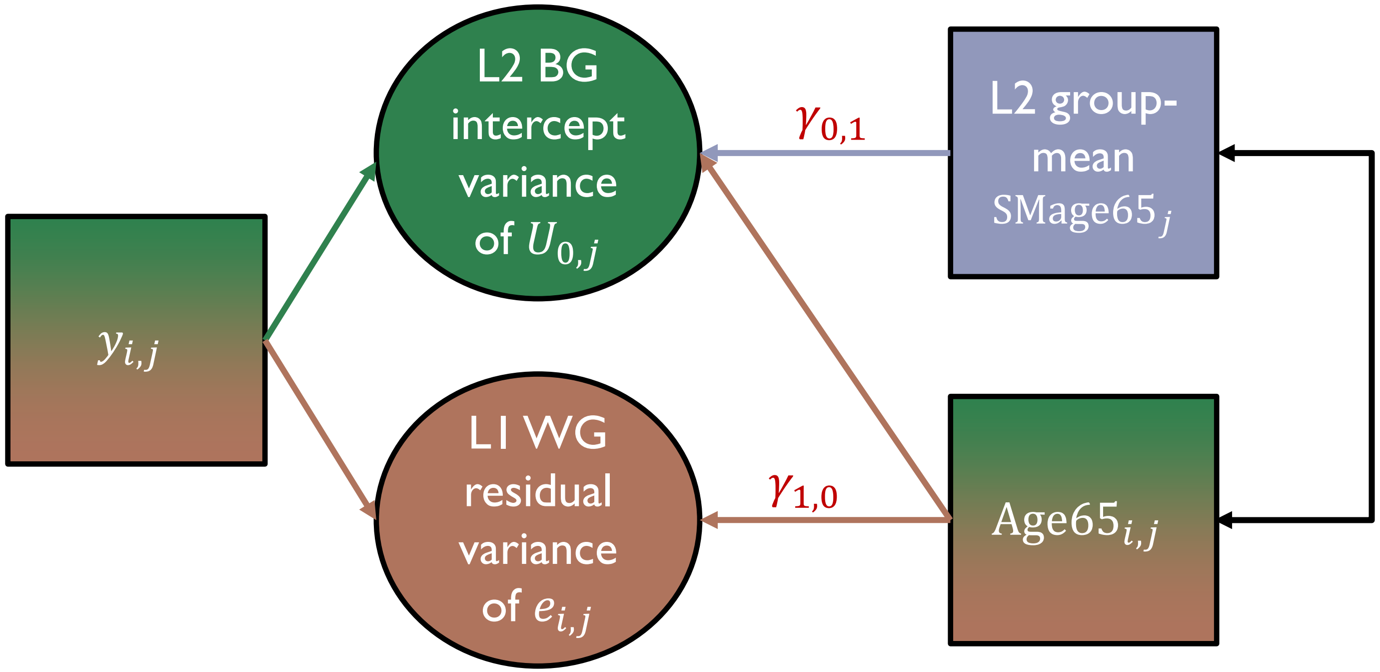
γ_{10} = <u>unique</u> within-group effect controlling for SMage65_j

γ_{01} = contextual effect = how the effect of SMage65_j differs from the effect of $\text{Age65}_{i,j}$ = <u>unique</u> between-group effect after controlling for $\text{Age65}_{i,j}$

Disaggregating Level-1 from Level-2

Model-based partitioning of $y_{i,j}$ outcome variance

No partitioning of $\text{Age65}_{i,j}$, but group-mean age $(\overline{\text{Age}_j} - 65)$ is included in the model to statistically remove the shared variance at level-2 from level-1 predictor



Because $\text{Age65}_{i,j}$ still has BG variance, it still carries some non-zero BG effect.
We "statistically control" for that level-2 partition by including SMage65_j.

RyanWalters@creighton.edu

Why We Require Statistical Control

- In grand-mean-centering, the level-1 variable retains level-1 and level-2 variability
  - We will see that $\text{Age65}_{i,j}$ is correlated with SMage65_j

- Consider a garden-variety linear regression model…

$$y_i = \beta_0 + \beta_1(X_{1,i}) + \beta_2(X_{2,i}) + e_i$$

  - If $X_{1,i}$ and $X_{2,i}$ are correlated then…
  - β_1 is the unique effect of $X_{1,i}$ after controlling for $X_{2,i}$
  - β_2 is the unique effect of $X_{2,i}$ after controlling for $X_{1,i}$

- Grand-mean-centering provides the within-group and contextual effects
  - WG: fixed $\text{Age65}_{i,j}$ effect is the unique level-1 effect after controlling for level-2 SMage65_j
  - Contextual: fixed SMage65_j effect is the unique level-2 effect after controlling for level-1 $\text{Age65}_{i,j}$

Grand-mean-centering: Three Effects

- Grand-mean-centering is likely more useful for clustered data as it can directly provide level-2 contextual effects

- Effects estimated directly by grand-mean-centering
  - Level-1, within-group: $\gamma_{1,0}$
  - Level-2, contextual: $\gamma_{0,1}$

- Effects not estimated directly by grand-mean-centering
  - Level-2, between-group: $\gamma_{0,1}$
  - Between = contextual + within = $\gamma_{0,1} + \gamma_{1,0}$

- Recommend using *lincom*, *ESTIMATE*, *TEST*, or *contrast1D* statements to request the missing third effect
  - Alternatively, you could use group-mean-centering to get the level-2, between-group effect directly (sit tight)

Example Time!

Example - Grand-mean-centering.pdf

Example - Grand-mean-centering.xlsx

(Interim Analyses for Age, Models 3a-3b)

Applied Multilevel Models

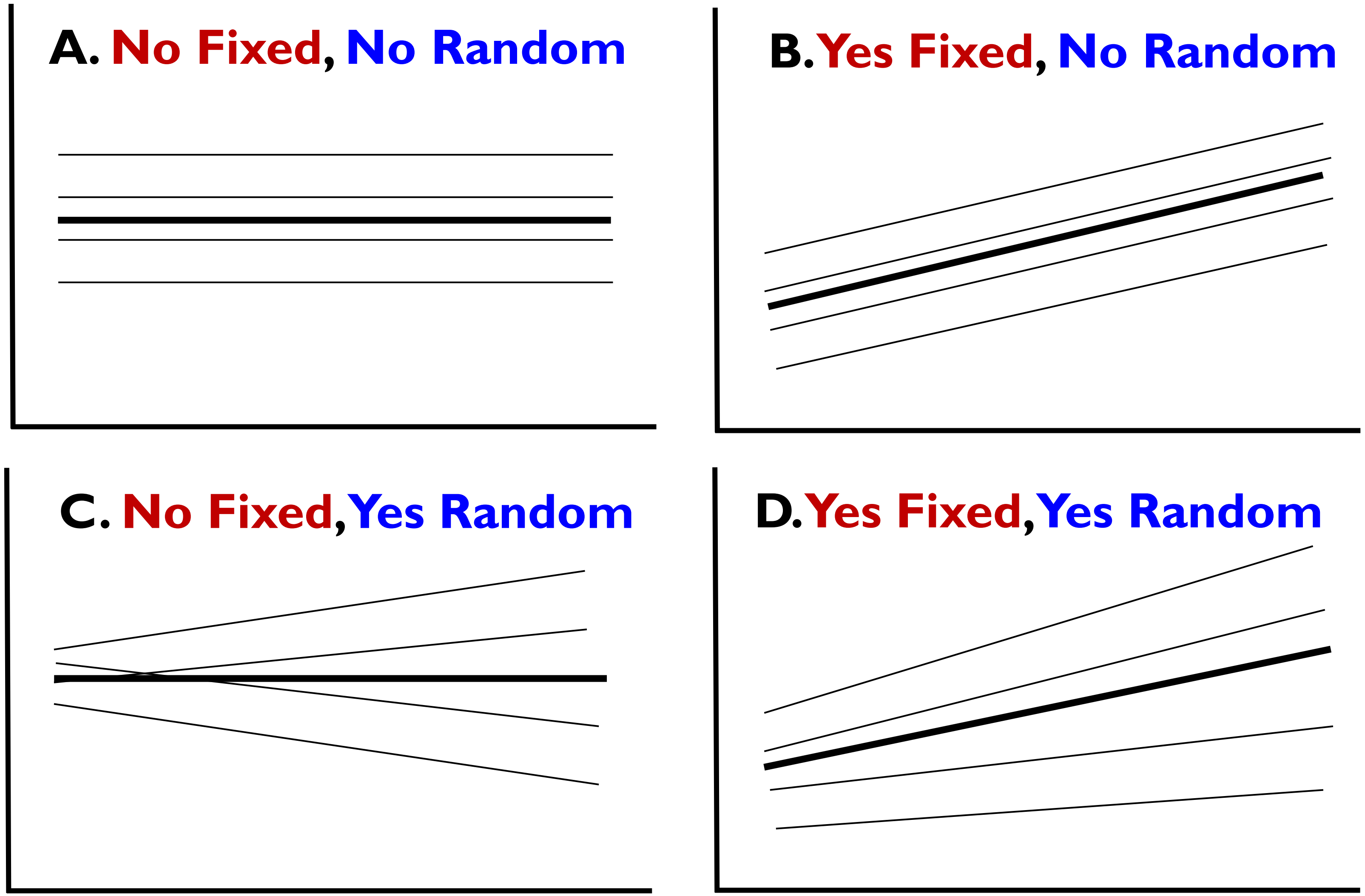
Part 9 of 12:

Random Effects of Level-1, Person-level Predictors

Fixed vs. Random Effects

- There are two questions specific to the effect of a level-1 predictor
- Question 1: is there an effect on average?
  - Non-flat slope
  - Significant fixed effect
- Question 2: does this average effect adequately describe every group in the sample?
  - Do groups need their own slope?
  - Significant random slope effect

Fixed and Random Effects



*Thick black line is the fixed effect. Thin black lines are group-specific effects.

Random Level-1 Effects

- Level 1

$$y_{i,j} = \beta_{0,j} + \beta_{1,j}(\text{Age65}_{i,j}) + e_{i,j}$$

Residual = person-specific deviation from their group's predicted outcome; variance = σ_e^2

Fixed intercept = predicted outcome when $\text{Age65}_{i,j}$ and SMage65_j = 0.

$\gamma_{0,1}$ = contextual effect = the unique between-group effect after controlling for $\text{Age65}_{i,j}$

- Level 2

$$\beta_{0,j} = \gamma_{0,0} + \gamma_{0,1}(\text{SMage65}_j) + U_{0,j}$$
$$\beta_{1,j} = \gamma_{1,0} + U_{1,j}$$

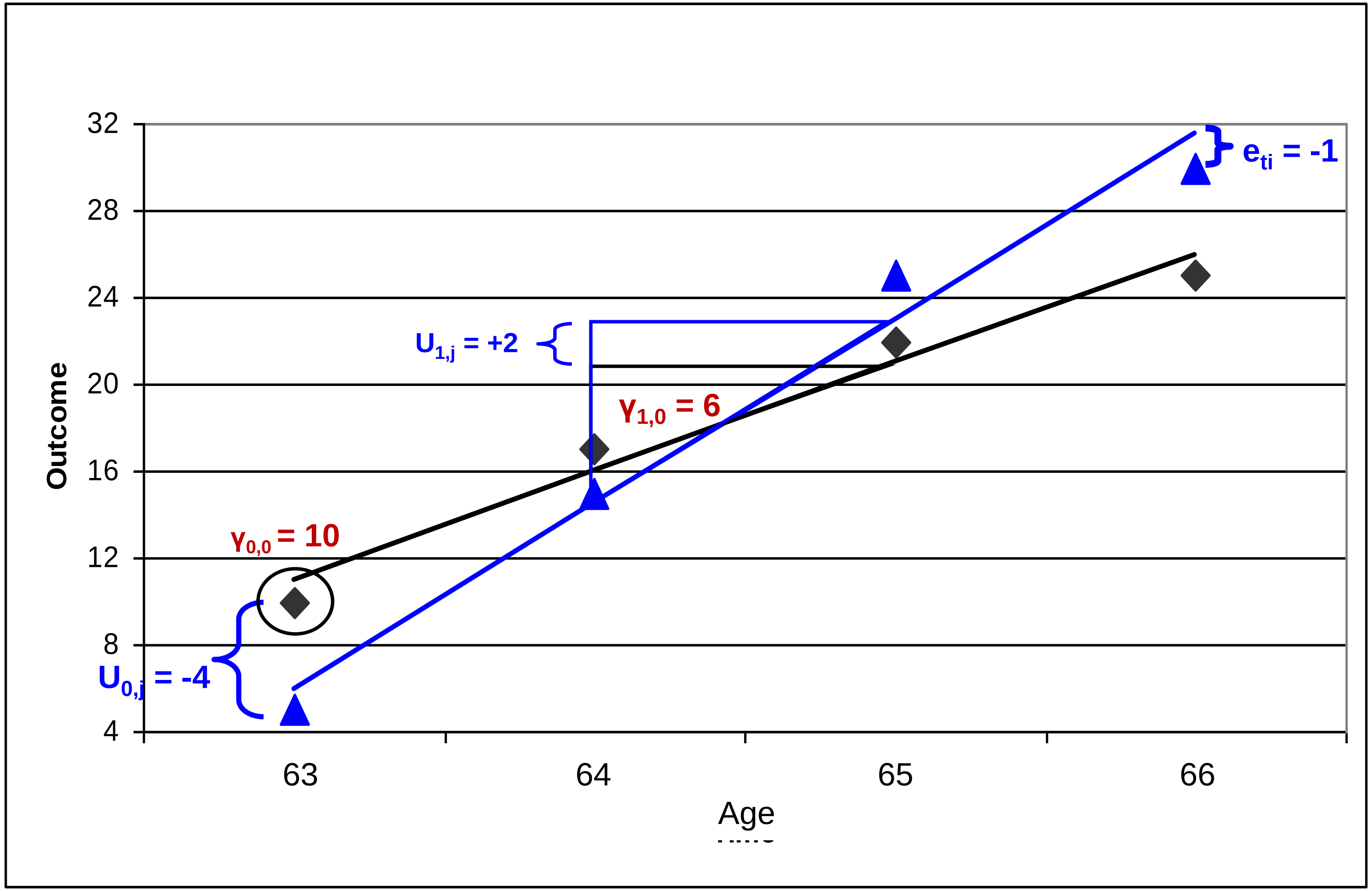
Random intercept = group-specific deviation from fixed intercept specifically at age 65; variance = $\tau_{U_0}^2$

Random within-group age slope = group-specific deviation from fixed within-group age slope; variance = $\tau_{U_1}^2$

$\gamma_{1,0}$ = the unique within-group effect after controlling for SMage65_j

There now also exists the covariance between the random intercept and random slope τ_{U_0,U_1}

Random Effect of Level-1 Age



More Random Effect Commentary

- Random slope variances are placed in the **G**-matrix
  - Generally, the **G**-matrix is <u>unstructured</u> meaning that every random effect variance and covariance is estimated

$$\mathbf{G} = \begin{bmatrix} \tau_{U_0}^2 & \tau_{U_1,U_2} \\ \tau_{U_2,U_1} & \tau_{U_1}^2 \end{bmatrix}$$

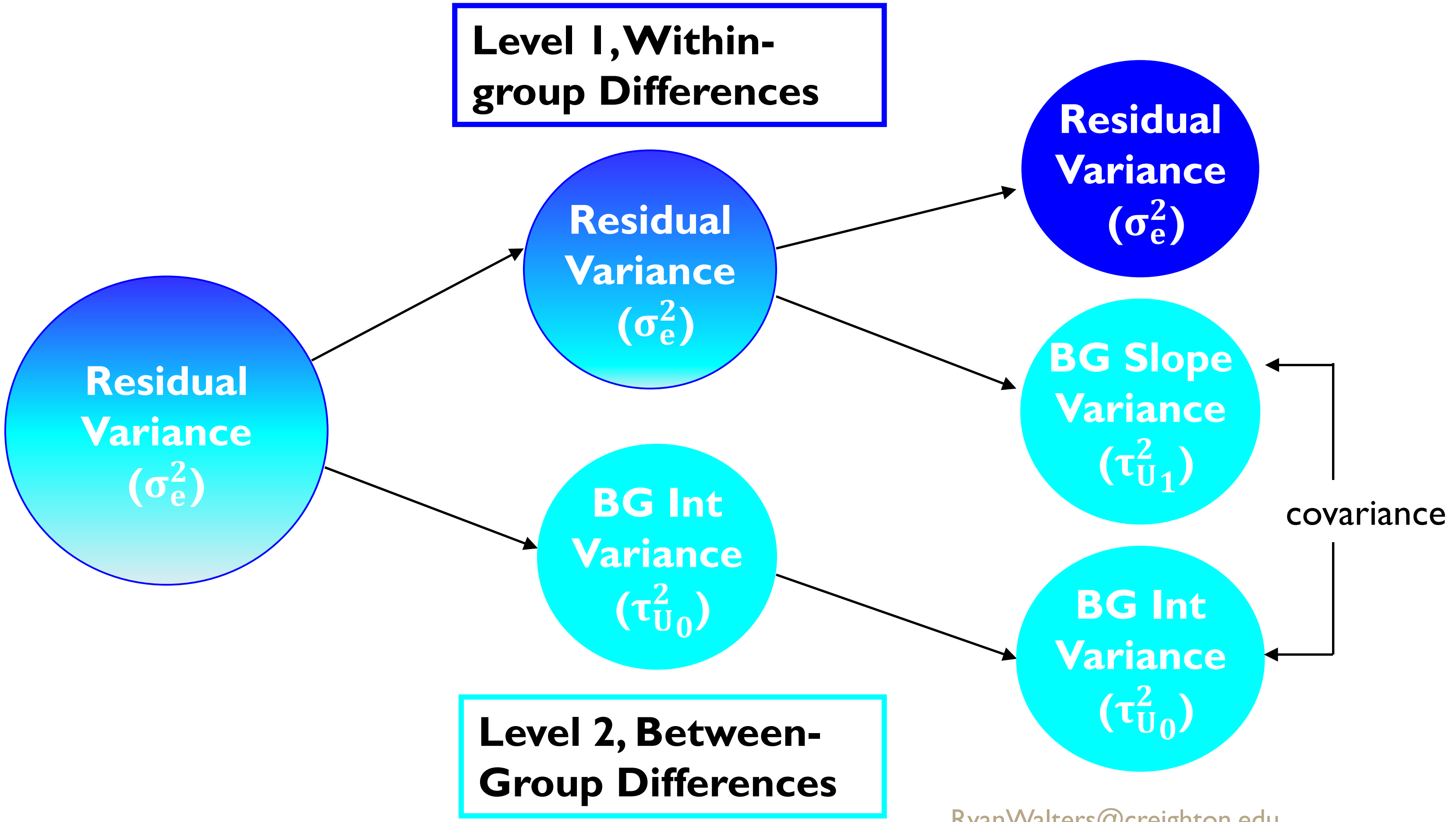
- The addition of new random effects is tested via model comparison via likelihood ratio test
  - Let us count the variance components…
    - Random intercept model: 1 (random intercept)
    - Random age slope model: 3 (random intercept, random slope, covariance)
  - Degrees of freedom = 3 – 1 = 2
  - If in REML, make sure both models have <u>the same</u> predictors variables

- We can quantify random slope variances in terms of the actual outcome using the 95% random effects confidence interval

- After including random slopes, we have to reset pseudo-R^2 calculations because…

How a Multilevel Model Handles Dependency

- Random slopes are partitioned out of residual variance and…
- The random intercept variance is now conditional the 0 value for the predictor with the random slope



More Commentary on Random Effects

- Before I begin, please suspend your disbelief. Okay…

- Random effects under grand-mean-centering have a hiccup
  - The random effect of level-1 age is a <u>convergence effect</u> even if we disaggregated that predictor's level-1 and level-2 fixed effects
  - There is forthcoming methodological literature to support this with algebra and examples via simulation

- This issue is conceptually identical to the fixed effects
  - We should "technically" also include the level-2 partition as a random effect to disaggregate the level-1 random effect
  - <u>Unique</u> random slope variance after controlling for level 2

- That said, in a two-level model, there is no higher-level for any level-2 effect to vary across
  - I have never gotten this model to actually estimate…

Example Time!

Example - Grand-mean-centering.pdf

Example - Grand-mean-centering.xlsx

(Models 3c-3d)

Applied Multilevel Models

Part 10 of 12:
Explaining Random Slope Variance

Explaining Random Slope Variance

- Recall that random effects represent level-2 between-group differences; thus, random effect variances can only be explained by level-2 variables

- Reconsider fixed effects of level-2 predictors by themselves
  - Level-2 main effects explain level-2 random intercept variance ($\tau_{U_0}^2$)
  - Level-2 interactions explain level-2 random intercept variance ($\tau_{U_0}^2$)

$$\text{Pseudo-}R^2_{\tau_{U_0}^2} = \frac{\text{random intercept variance}_1 - \text{random intercept variance}_2}{\text{random intercept variance}_1}$$

- Random slope variances are explained by fixed cross-level interaction effects
  - An interaction between a level-2 variable and the random level-1 variable
  - i.e., a BG*WG interaction

$$\text{Pseudo-}R^2_{\tau_{U_1}^2} = \frac{\text{random slope variance}_1 - \text{random slope variance}_2}{\text{random slope variance}_1}$$

The Joy of Interactions Involving Level-1 Predictors

- Grand-mean-centering: the level-1 variable contains within- and between-group variability
  - Just like main effects, interaction effects have to take this fact into account

- Consider how level-2 state unemployment moderates level-1 person age and vice versa
  - SMue2_j*$\text{Age65}_{i,j}$: does the WG age effect differ by state unemployment?
  - $\text{SMue2}_j * \text{Age65}_{i,j}$: does the state unemployment effect differ by person age?

- Say your focus is on the cross-level interaction $\text{Age65}_{i,j}$*SMue2_j
  - It is <u>not okay</u> to omit SMage_j*SMue2_j
  - Remember, the WG and BG age effects are correlated
  - Although the level-1 effect of age ($\text{Age65}_{i,j}$) is <u>not</u> a convergence effect (because of SMage_j), the $\text{Age65}_{i,j}$*SMue2_j interaction <u>would be</u> a convergence effect!
  - SMue2_j*SMage_j: does the contextual age effect differ by state unemployment?
  - $\text{SMue2}_j * \text{SMage}_j$: does the state unemployment effect differ by contextual age?

Example Time!

Example - Grand-mean-centering.pdf

Example - Grand-mean-centering.xlsx

(Models 4a-4b)

Applied Multilevel Models

Part 11 of 12:

Group-mean-centering

Group-mean-centering

- Decomposes the level-1 predictor into two variables that directly represent either the between-group or the within-group sources of variation

- Consider people clustered in states and level-1 predictor $\text{Age65}_{i,j}$

- Level-2, group-mean predictor (same as grand-mean-centering)
  - SMage65_{j} = $\overline{\text{Age}}_{j} - 65$ = centered state-mean age
  - State-mean age is typically based on the sample data
  - As usual, SMage_{j} was centered to ensure meaningful 0

- Level-1, within-group predictor (big difference from grand-mean-centering)
  - $\text{WSage}_{i,j}$ = $\text{Age65}_{i,j} - \text{SMage}_{j}$ = deviation from state's mean age
  - $\text{WSage}_{i,j}$ is centered at a variable, not a constant
  - Positive $\text{WSage}_{i,j}$ = person is older than state mean
  - Negative $\text{WSage}_{i,j}$ = person is younger state mean

Group-Mean-Centering

- Within- and between-group effects via separate predictors
  - $\text{Age65}_{i,j}$ is group-mean-centered into level-1 $\text{WSage}_{i,j}$ with SMage65_j at level-2

- Level 1

  $y_{t,i} = \beta_{0,i} + \beta_{1,i}(\text{WSage}_{i,j}) + e_{t,i}$

- Level 2

  $\beta_{0,i} = \gamma_{00} + \gamma_{01}(\text{SMage65}_j) + U_{0,i}$

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

γ_{10} = within-group main effect of being older than others in the same state

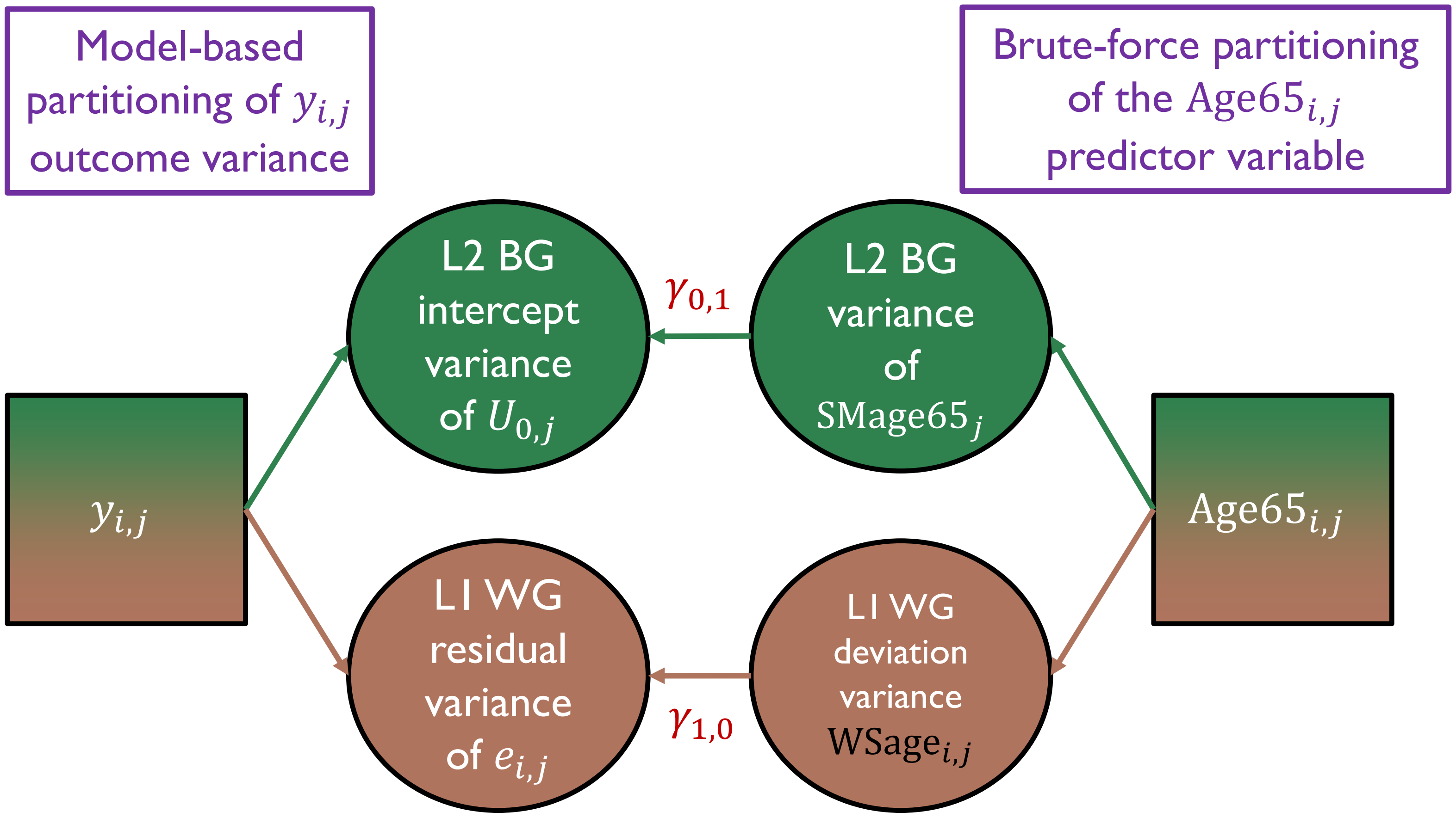
γ_{01} = between-group main effect of living in an state in which people are older

$\text{WSage}_{i,j} = \text{Age65}_{i,j} - \text{SMage65}_j$
Only within-group variation

$\text{SMage65}_j = \overline{\text{Age}}_j - 65$
Only between-group variation

Because $\text{WSage}_{i,j}$ and SMage65_j are uncorrelated, each gets the total effect for its level

Disaggregating Level-1 from Level-2



$\text{WSage}_{i,j}$ contains zero BG variance so there is no statistical control.
It is more difficult to make interpretational mistakes with group-mean-centering

No Statistical Control Required

- In group-mean-centering, the level-1 variable is only level-1
  - We will see that $\mathrm{WSage}_{i,j}$ has zero correlation with $\mathrm{SMage65}_j$
  - In fact, $\mathrm{WSage}_{i,j}$ is uncorrelated with all level-2 variables!

- Let us return to the garden-variety linear regression model…

$$y_i = \beta_0 + \beta_1\left(X_{1,i}\right) + \beta_2\left(X_{2,i}\right) + e_i$$

  - If $X_{1,i}$ and $X_{2,i}$ are uncorrelated there is no statistical control, so…
  - β_1 is all the relationship between $X_{1,i}$ and y_i
  - β_2 is all the relationship between $X_{2,i}$ and y_i

- Group-mean-centering provides within- and between-group effects
  - WG: the fixed effect of $\mathrm{WSage}_{i,j}$ is a level-1 effect
  - BG: the fixed effect of $\mathrm{SMage65}_j$ is a level-2 effect

Group-mean-centering: Three Effects

- Effects given directly by the model
  - Level-1, within-group: $\gamma_{1,0}$
  - Level-2, between-group: $\gamma_{0,1}$
- Effects not given directly by the model
  - Level-2, contextual: $\gamma_{0,1}$
  - Contextual = between – within = $\gamma_{0,1} - \gamma_{1,0}$
- Recommend using *lincom*, *ESTIMATE*, *TEST*, or *contrast1D* statements to request the missing third effect
  - Alternatively, you could use grand-mean-centering to get the level-2, contextual effect directly

Random Level-1 Effects

- Level 1

$$y_{i,j} = \beta_{0,j} + \beta_{1,j}(\text{WSage65}_{i,j}) + e_{i,j}$$

Residual = person-specific deviation from the group's predicted outcome; variance = σ_e^2

Fixed intercept = predicted outcome when $\text{WSage65}_{i,j}$ and SMage65_j = 0

$\gamma_{0,1}$ = between-group main effect of living in a state in which people are older

- Level 2

$$\beta_{0,j} = \gamma_{0,0} + \gamma_{0,1}(\text{SMage65}_j) + U_{0,j}$$
$$\beta_{1,j} = \gamma_{1,0} + U_{1,j}$$

Random intercept = group-specific deviation from fixed intercept specifically at age 65; variance = $\tau_{U_0}^2$

Random within-group age slope = group-specific deviation from fixed within-group age slope; variance = $\tau_{U_1}^2$

$\gamma_{1,0}$ = within-group of being older than people in their own state

There now also exists the covariance between the random intercept and random slope τ_{U_0,U_1}

Random Slopes

- Random slope variances are still placed in an unstructured **G**-matrix

$$\mathbf{G} = \begin{bmatrix} \tau^2_{U_0} & \tau_{U_1,U_2} \\ \tau_{U_2,U_1} & \tau^2_{U_1} \end{bmatrix}$$

- No need to consider the random effect of that level-2 partition!
- New random effects are tested via likelihood ratio test
  - If in REML, make sure both models have <u>the same</u> predictors variables
- We quantify random slope variances in terms of the actual outcome using the 95% random effects confidence interval
- With random slopes, we reset pseudo-R^2 calculations

Explaining Random Slope Variance

- Recall that random effects represent level-2 between-group differences; thus, random effect variances can only be explained by level-2 variables

- Reconsider fixed effects of level-2 predictors by themselves
  - Level-2 main effects explain level-2 random intercept variance ($\tau_{U_0}^2$)
  - Level-2 interactions explain level-2 random intercept variance ($\tau_{U_0}^2$)

$$\text{Pseudo}-R^2_{\tau^2_{U_0}} = \frac{\text{random intercept variance}_1 - \text{random intercept variance}_2}{\text{random intercept variance}_1}$$

- Random slope variances are explained by fixed cross-level interaction effects
  - An interaction between a level-2 variable and the random level-1 variable
  - i.e., a BG*WG interaction

$$\text{Pseudo}-R^2_{\tau^2_{U_1}} = \frac{\text{random slope variance}_1 - \text{random slope variance}_2}{\text{random slope variance}_1}$$

Interactions with Level-1 Predictors

- Under group-mean-centering, the level-1 variable only contains within-group variability, so interactions are (relatively) simpler

- Consider how level-2 state unemployment ($\mathrm{SMue2}_j$) moderates level-1 person age ($\mathrm{WSage65}_{i,j}$) and vice versa
  - $\mathrm{SMue2}_j$*$\mathrm{WSage65}_{i,j}$: does the WG age effect differ by state unemployment?
  - $\mathrm{SMue2}_j * \mathrm{WSage65}_{i,j}$: does the state unemployment effect differ by person age?

- Say your focus is on the cross-level interaction $\mathrm{WSage65}_{i,j}$*$\mathrm{SMue2}_j$
  - It is okay to omit SMage_j*$\mathrm{SMue2}_j$, but that is kind of weird because…
  - You would be saying that $\mathrm{SMue2}_j$ moderates level-1 age, but not level-2 age even though they were created from the same variable
  - $\mathrm{SMue2}_j$*SMage_j: does the BG age effect differ by state unemployment?
  - $\mathrm{SMue2}_j * \mathrm{SMage}_j$: does the state unemployment effect differ by BG age?

Example Time!

Example - Group-mean-centering.pdf

Example - Group-mean-centering.xlsx

(Models 3a-4b)

(Note: Models 1-2b are identical to group-mean-centering)

Applied Multilevel Models

Part 12 of 12:
Overarching Summary and Model Building Advice

Overarching Summary

- Multilevel models for clustered data come in two varieties
  - Empty vs. conditional
- Level-1 predictors carry at least two effects in a two-level model
  - Level-2, BG: some groups are higher/lower than other groups (fixed only)
  - Level-1, WG: some people are higher or lower than others in their group
    - Can be fixed or random
- BG and WG effects almost always need to be represented by two or more model parameters using either…
  - Group-mean-centering asking whether WG ≠ 0? BG ≠ 0?
  - Grand-mean-centering asking whether WG ≠ 0? BG ≠ WG?
- Grand-MC makes more sense for clustered data given interest often lies in contextual effect

Model Building Strategies

- Calculate that ICC
  - Calculate the 95% RECI around that fixed intercept

- Include level-2, group-level predictors
  - Use p-values to determine statistical significance
  - Calculate pseudo-R^2 and/or total R^2

- Include level-1, person-level predictors
  - Calculate the ICC for that predictor
  - If ICC > 0, disaggregate predictor fixed effects via grand-mean-centering or group-mean-centering
  - Use p-values to determine statistical significance
  - Make sure to keep interpretations straight (WG vs. BG vs. contextual)

- Evaluate random slopes for level-1, person-level predictors
  - Use the likelihood ratio test to determine statistical significance
  - Calculate the 95% RECI for each random slope variance
  - Include cross-level interactions to explain random slope variance
  - Calculate pseudo-R^2 and/or total R^2

Model-Building Strategies – Part Deux

- This workshop used a bottom-up model-building approach
- It may be helpful to examine predictor effects in separate models first
  - e.g., does age matter at all
- Then combine predictor effects in layers in order to examine their unique contribution (and interactions)
  - e.g., does age still matter after considering biologic sex?
- 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
  - You may not end up in the same place give differential order of predictor entry