Two-Level Models for Clustered* Data

- Today's Class:
  - 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
  - Nested vs. Crossed Groups Designs

* *Clustering = Nesting = Grouping…*

MLM for Clustered Data

- So far we've built models to account for dependency created by repeated measures (time within person, or trials within persons crossed with items)

- Now we examine two-level models for more general examples of nesting/clustering/grouping:
  - Students within schools, athletes within teams, patients within doctors
  - Siblings within families, partners within dyads

- 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:

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

Fixed Intercept =grand mean (because no predictors yet)

Random Intercept = group-specific deviation from predicted intercept

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

3 Total Parameters:

Model for the Means (1):

- Fixed Intercept γ_{00}

Model for the Variance (2):

- Level-1 Variance of e_{ij} → σ_e^2
- Level-2 Variance of U_{0j} → $\tau_{U_0}^2$

Composite equation:

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

Matrices in a Random Intercept Model

RI and DIAG: 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 = $\tau^2_{U_0}$ / ($\tau^2_{U_0}$ + σ^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)

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

$$= \frac{\boldsymbol{\tau}^2_{\mathbf{U_0}}}{\boldsymbol{\tau}^2_{\mathbf{U_0}} + \boldsymbol{\sigma}^2_{\mathbf{e}}}$$

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

- ICC = Proportion of total variance that is between <u>groups</u>
- ICC = Average correlation among persons from same <u>group</u>
- ICC is a standardized way of expressing how much we need to worry about *dependency due to <u>group</u> mean differences* **(i.e., ICC is an effect size for *constant* <u>group</u> 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 using a given sampling design to the variance using 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 → $\text{N}_{\text{effective}} = \frac{\text{\# Total Observations}}{\text{Design Effect}}$

- As ICC goes UP and cluster size goes UP, the effective sample size goes DOWN
  - See Snijders & Bosker (2012) for more info and for a modified formula that takes unequal group sizes into account

Design Effects in 2-Level Nesting

- Design Effect = $1 + [(n-1) * \text{ICC}]$
- Effective sample size → $\text{N}_{\text{effective}} = \dfrac{\text{\# Total Observations}}{\text{Design Effect}}$

- 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)

Level-1 (Person-Level) Predictors

- Modeling of level-1 predictors is complicated (and usually done incorrectly) because **each level-1 predictor is usually really 2 predictor variables** (each with their own effect), **not 1**

- Example: Student SES when students are clustered in schools
  - Some kids have more money than other kids in their school:
    - **WG variation in SES** *(represented directly as deviation from school mean)*
  - Some schools have more money than other schools:
    - **BG variation in SES** *(represented as school mean SES or via external info)*

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

Between-Group vs. Within-Group Effects

- Between-group and within-group effects in SAME direction
  - SES → Achievement?
    - **BG: Schools with more money than other schools may have greater mean achievement than schools with less money**
    - **WG: Kids with more money than other kids in their school may have greater achievement than other kids in their school (regardless of school mean SES)**

- Between-group and within-group effects in OPPOSITE directions
  - Body mass → life expectancy in animals (Curran and Bauer, 2011)?
    - **BG: Larger species tend to have longer life expectancies than smaller species (e.g., whales live longer than cows, cows live longer than ducks)**
    - **WG: Within a species, relatively bigger animals have shorter life expectancy (e.g., over-weight ducks die sooner than healthy-weight ducks)**

- Variables have **different meanings** and **different scales** across levels (so "one-unit" effects will rarely be the same across levels)!

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 $\mathrm{x_{ij}}$ from your sample (labeled as $\mathbf{GMx_j}$ or $\bar{\mathbf{X}}_\mathbf{j}$), centered at a constant so that 0 is a meaningful value
  - Use **actual** group mean of level-1 $\mathrm{x_{ij}}$ 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: $\mathbf{WGx_{ij}} = \mathbf{x_{ij}} - \bar{\mathbf{X}}_\mathbf{j}$
  - Level-1 Grand-MC → center at a CONSTANT: $\mathbf{L1x_{ij}} = \mathbf{x_{ij}} - \boldsymbol{C}$
    - Use $\mathrm{L1x_{ij}}$ 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 (σ^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

γ_{10} = WG main effect of having more x_{ij} than others in your group

γ_{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 (σ^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 <u>directly</u>:**

- **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 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 "**<u>grand-mean-centering</u>**" for level-1 x_{ij} instead: $\mathbf{L1x_{ij} = x_{ij} - \boldsymbol{C}}$

    → **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…

WRONG WAY: Clustered Data Model with x_{ij} represented at Level 1 Only: → WG and BG Effects are <u>**Smushed Together**</u>

x_{ij} is grand-mean-centered into $L1x_{ij}$, <u>WITHOUT</u> 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: $\beta_{0j} = \gamma_{00} + U_{0j}$

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

γ_{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

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

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 differer** (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 $L1x_{ij}$, <u>WITH</u> GMx_j at L2:

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

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

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

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

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

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

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

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

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

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

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

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

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

Effect	Group-MC	Grand-MC
Intercept	γ_{00}	γ_{00}
WG Effect	γ_{10}	γ_{10}
Contextual	$\gamma_{01} - \gamma_{10}$	γ_{01}
BG Effect	γ_{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 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 ($\tau^2_{U_0}$)?
  - 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 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 (σ^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 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 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 σ^2_e decreases, $\tau^2_{U_0}$ increases

The Joy of Interactions Involving Level-1 Predictors

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

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

- Grand-Mean-Centering:
  - $\mathbf{L1x_{ij}} * \mathbf{Type_j}$ → Does the WG motivation effect differ between business types?
  - $\mathbf{GMx_j} * \mathbf{Type_j}$ → Does the *contextual* motivation effect differ b/t business types?
    - 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 $\mathbf{GMx_j}$, the interaction of $\mathbf{L1x_{ij}} * \mathbf{Type_j}$ would still be smushed

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

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

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

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

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

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

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

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

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

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

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:
  - $\mathbf{WGx_{ij}} * \mathbf{GMx_j}$ → Does the WG motivation effect differ by group motivation?
  - $\mathbf{GMx_j} * \mathbf{GMx_j}$ → Does the BG motivation effect differ by group motivation?
    - Moderation of total group motivation effect (not controlling for individual motivation)
    - If forgotten, then $\mathbf{GMx_j}$ moderates the motivation effect only at level 1 (WG, not BG)

- Grand-Mean-Centering:
  - $\mathbf{L1x_{ij}} * \mathbf{GMx_j}$ → Does the WG motivation effect differ by group motivation?
  - $\mathbf{GMx_j} * \mathbf{GMx_j}$ → 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 $\mathbf{GMx_j}$, the interaction of $\mathbf{L1x_{ij}} * \mathbf{GMx_j}$ 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: $\gamma_{00} = \gamma_{00}$

WG Effect: $\gamma_{10} = \gamma_{10}$

BG Effect: $\gamma_{01} = \gamma_{01} + \gamma_{10}$

BG² Effect: $\gamma_{02} = \gamma_{02} + \gamma_{11}$

Contextual: $\gamma_{01} = \gamma_{01} - \gamma_{10}$

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

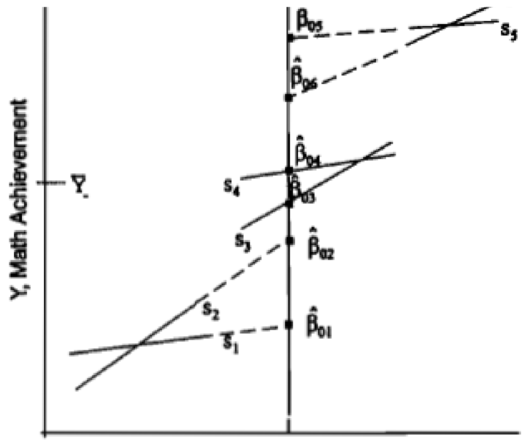
- **Random intercepts** mean different things under each model:
  - **Group-MC** → Group differences at $\mathbf{WGx_{ij} = 0}$ (that every group has)
  - **Grand-MC** → Group differences at $\mathbf{L1x_{ij} = 0}$ (that not every group will have)

- **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)

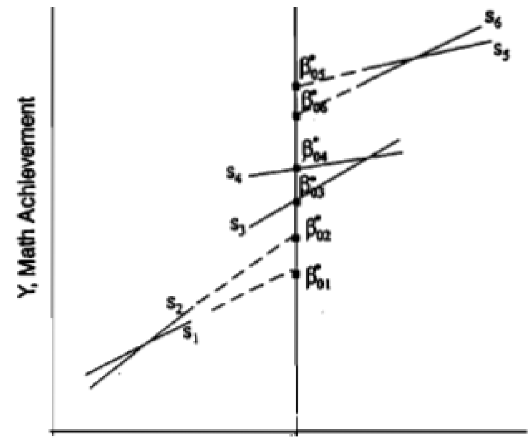
Bias in Random Slope Variance

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

MLM for Clustered Data: Summary

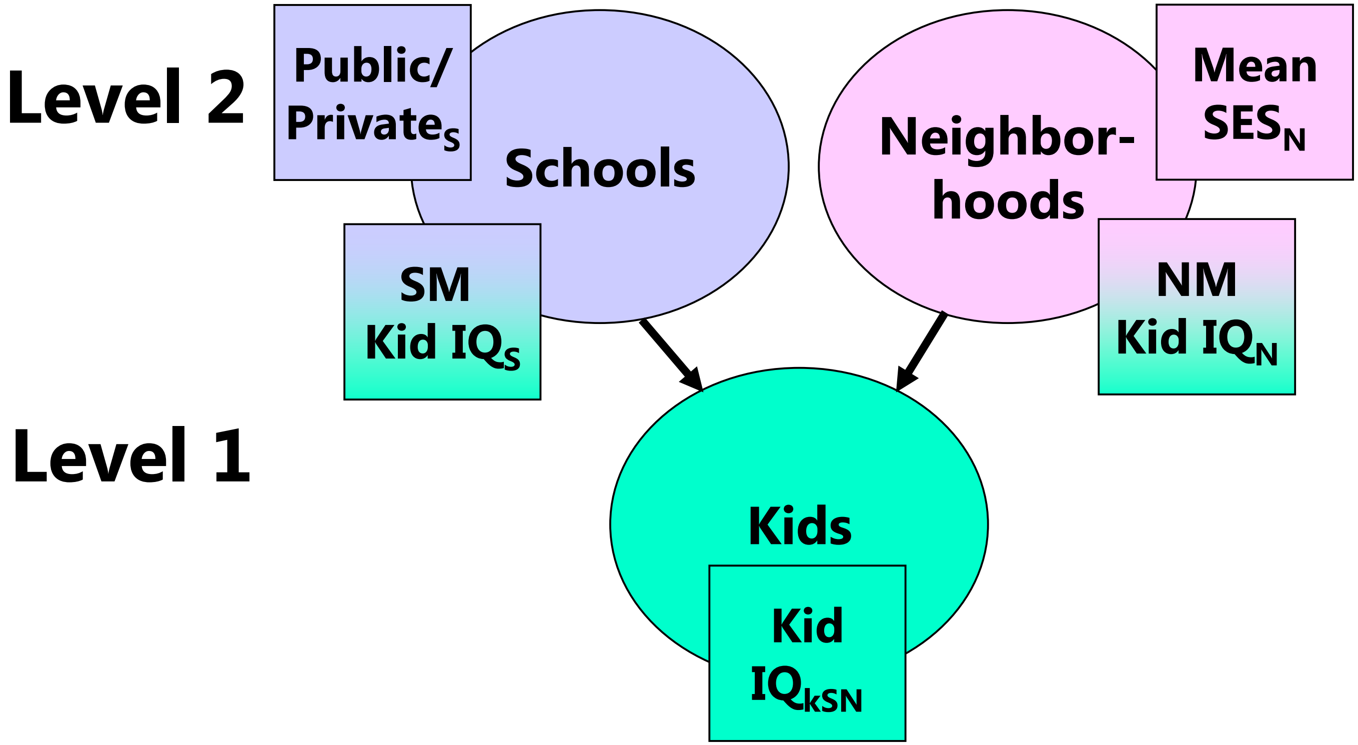
- Models now come in only two kinds: "empty" and "conditional"
  - The lack of a comparable dimension to "time" simplifies things greatly!
- L2 = Between-Group, L1 = Within-Group (between-person)
  - Level-2 predictors are group variables: can have fixed or systematically varying effects (but not random 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

More Complex Multilevel Designs

- Multilevel models are specified based on the relevant dimensions by which observations differ each other, and how the units are organized
- Two-level models have at least two piles of variance, in which level-1 units are nested within level-2 units:
  - Longitudinal Data: Time nested within Persons
  - Students nested within Teachers
- Three-level models have at least three piles of variance, in which level-2 units are nested within level-3 units:
  - Time nested within Persons within Families
  - Student nested within Teachers within Schools
- In other designs, multiple sources of systematic variation may be present, but the sampling may be crossed instead…
  - Same idea as crossed random effects (i.e., as for persons and items), but these are known as “cross-classified” models in the clustered data world
  - Here are a few examples on when this might happen…

Kids, Schools, and Neighborhoods

- Kids are nested within schools AND within neighborhoods
- Not all kids from same neighborhood live in same school, so schools and neighborhoods are crossed at level 2
- Can include predictors for each source of variation



Kids, Schools, and Neighborhoods

$y_{kSN} = \gamma_{000}$ → *fixed intercept (all x's = 0)*

$+ \gamma_{010}(\text{Private}_S) + \gamma_{020}(\text{SMIQ}_S)$ → *school effects*

$+ \gamma_{001}(\text{SES}_N) + \gamma_{002}(\text{NMIQ}_N)$ → *neighborhood effects*

$+ \gamma_{100}(\text{KidIQ}_{kSN})$ → *kid effects*

$+ U_{0S0}$ → *random effect of school*

$+ U_{00N}$ → *random effect of neighborhood*

$+ e_{kSN}$ → *residual kid-to-kid variation*

Time (t), Students (s), and Classes (c)

- Students are nested within Classes at each occasion…
- But if students move into different classes across time…
  - Time at level 1 is nested within Student AND within Classes
  - Student is crossed with Class at level 2
- How to model a time-varying random classroom effect?
  - This is the basis of so-called "value-added models"

- (At least) Two options via fixed or random effects:

  - Acute effect: Effect for class operates only when kids are in the class
    - e.g., Class effect ← teacher bias
    - Once a student is out of the class, class effect is no longer present

  - Transfer effect: Effect for class operates now and in the future…
    - e.g., Class effect ← differential learning
    - Effect stays with the student in the future (i.e., a "layered" value-added model)

Time (t), Students (s), and Classes (c)

- Custom-built intercepts for time-varying effects of classes
  - An intercept is usually a column of 1's, but ours will be 0's and 1's to serve as switches that turn on/off class effects

				Per-Year Class ID (−99 = missing)			Intercepts for Acute Effects			Intercepts for Transfer Effects		
Student ID	Class ID	Grade	Year	Year 0 Class	Year 1 Class	Year 2 Class	Year 0 Intercept	Year 1 Intercept	Year 2 Intercept	Year 0 Effect	Year 1 Effect	Year 2 Effect
101	1	3	0	1	−99	43	1	0	0	1	0	0
101	−99	4	1	1	−99	43	0	0	0	0	0	0
101	43	5	2	1	−99	43	0	0	1	1	0	1
102	3	3	0	3	21	42	1	0	0	1	0	0
102	21	4	1	3	21	42	0	1	0	1	1	0
102	42	5	2	3	21	42	0	0	1	1	1	1

Time (t), Students (s), and Classes (c)

- Hoffman (2015) Equation 11.3: fixed effects model for class as a categorical time-varying predictor:
  - Allows for control of classes only….

$$
\begin{aligned}
\text{Effort}_{tsc} = {} & \gamma_{000} + \gamma_{100}\left(\text{Year01}_{tsc}\right) + \gamma_{200}\left(\text{Year12}_{tsc}\right) + U_{0s0} + e_{tsc} \\
& + \gamma_{001}^{0}\left(\text{Class1}_{c}\right)\left(\text{Int0}_{tsc}\right) + \gamma_{002}^{0}\left(\text{Class2}_{c}\right)\left(\text{Int0}_{tsc}\right)\cdots + \gamma_{00C}^{0}\left(\text{ClassC}_{c}\right)\left(\text{Int0}_{tsc}\right) \\
& + \gamma_{001}^{1}\left(\text{Class1}_{c}\right)\left(\text{Int1}_{tsc}\right) + \gamma_{002}^{1}\left(\text{Class2}_{c}\right)\left(\text{Int1}_{tsc}\right)\cdots + \gamma_{00C}^{1}\left(\text{ClassC}_{c}\right)\left(\text{Int1}_{tsc}\right) \\
& + \gamma_{001}^{2}\left(\text{Class1}_{c}\right)\left(\text{Int2}_{tsc}\right) + \gamma_{002}^{2}\left(\text{Class2}_{c}\right)\left(\text{Int2}_{tsc}\right)\cdots + \gamma_{00C}^{2}\left(\text{ClassC}_{c}\right)\left(\text{Int2}_{tsc}\right)
\end{aligned}
$$

- Hoffman (2015) Equation 11.4: class as a random effects crossed with students at level 2:
  - Controls and models class-related variance so it can be predicted

$$
\begin{aligned}
\text{Effort}_{tsc} = {} & \gamma_{000} + \gamma_{100}\left(\text{Year01}_{tsc}\right) + \gamma_{200}\left(\text{Year12}_{tsc}\right) + U_{0s0} + e_{tsc} \\
& + U_{00c}^{0}\left(\text{Int0}_{tsc}\right) + U_{00c}^{1}\left(\text{Int1}_{tsc}\right) + U_{00c}^{2}\left(\text{Int2}_{tsc}\right)
\end{aligned}
$$

More on Cross-Classified Models

- In crossed models, lower-level predictors can have random slopes of over higher levels AND random slopes of the other crossed factor at the same level
  - Example: Kids, Schools, and Neighborhoods (data permitting)
    - Kid effects could vary over schools AND/OR neighborhoods
    - School effects could vary over neighborhoods (both level 2)
    - Neighborhood effects could vary over schools (both level 2)

- Concerns about smushing still apply over both level-2's
  - Separate contextual effects of kid predictors for schools and neighborhoods (e.g., after controlling for how smart you are, it matters incrementally whether you go to a smart school AND if you live in a neighborhood with smart kids)

Summary: Nested or Crossed Designs

- Dimensions of sampling can result in systematic differences (i.e., dependency) that needs to be accounted for in the model for the variances
  - Sometimes this dependency is from nested sampling
  - Sometimes this dependency is from crossed sampling

- Multilevel models that include crossed random effects (or cross-classified models):
  - Can address this dependency (statistical motivation)
  - Can quantify and predict the amount of variation due to each source (substantive motivation)
  - Can include simultaneous hypothesis tests pertaining to each source of variation (substantive motivation)