Lecture 7: Two-Level Models for Clustered Data

- **Modeling Dependency in Clustered Data**
- Grand-Mean-Centering vs. Group-Mean Centering in Multilevel Models for Clustered Data
- Effect Size via Pseudo-R^2
- Clustered Data Example in SAS, SPSS, and STATA

MLM for "Clustered" Data

Nesting/Clustering/Grouping as a Source of Correlation

- Up to this point, we've only talked about a special case of nesting due to repeated measures → time within person
- Now we examine two-level models for more general examples of nesting/clustering/grouping:
  - Child within school, student within team
  - Sibling within family, partner within couple
  - Employee within business, patient within doctor
- Residuals of people within groups are likely to be correlated due to group differences (e.g., purposeful grouping or shared experiences)
  - Dependency will make SEs of fixed effects too small (*p*-values too good)

Two Options to Deal with Grouping

Clustering/Nesting/Grouping as Fixed Effects

- Include (#groups-1) dummy codes for group membership in the model for the means → ***so group is not another "level"***
- Inference about ***specific*** differences between groups via fixed effects for group, but then you ***cannot*** include between-groups predictors
- Snijders & Bosker ch.4, p. 44 recommend if #groups < 10ish

Clustering/Nesting/Grouping as a Random Effect

- Estimate a ***variance*** for group differences in model for variances, such that ***group becomes another "level" of the model***
- Makes an inference about ***population*** of groups via random effects variances, for which you ***can*** include between-groups predictors
- Better option if #groups > 10ish

Multilevel Models for Clustered Data: 2-Level Empty Model

Longitudinal MLM

L1: $y_{ti} = \beta_{0i} + e_{ti}$

Clustered MLM

L1: $y_{ij} = \beta_{0j} + e_{ij}$

L2: $\beta_{0j} = \gamma_{00} + U_{0j}$

Fixed Intercept (Grand Mean)

Cluster Mean Deviation

Fixed Effects:

γ_{00} → grand mean intercept

Random Effects:

U_{0j} → mean deviation for cluster j

→ intercept variance of τ_{U0}^2

Error:

e_{ij} → deviation for person i

→ residual variance of σ_e^2

IntraClass Correlation for <u>2-Level Empty Model</u>

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

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

- Average correlation among level 1 observations within a level 2 unit (within-**GROUP** correlation over people)
- <u>Intercept variance</u> now represents:
  "Why don't all groups have the same mean of Y?"
- <u>Residual variance</u> now represents:
  "Why don't all people in the same group have the same Y?"

Effects of Clustering on Effective *N*

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

- **Design Effect** = 1 + (($n_{\text{lower level}}$ – 1) * ICC)
- $\mathbf{N_{effective}}$ = #Total observations / design effect

- As ICC goes UP and cluster size goes UP, effective sample size goes DOWN
- See Snijders & Bosker ch. 3, p. 22-24 for more info

Design Effects in 2-Level Nesting

Design Effect = 1 + (($n_{lower\ level}$ – 1) * ICC)

$N_{effective}$ = #Total observations / design effect

- 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)
- 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 the grouping?

- The 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 uniformly “safe” to ignore, not even 0, because…
- …Unconditional and conditional (after predictors) ICCs may differ
  - Reducing the residual variance often results in an increase in the random intercept variance, which then increases the conditional ICC
- So just do a multilevel analysis…
  - Even if “that’s not your question”… you still have to care that your data are clustered and model that dependency appropriately because of:
    - Effect of clustering on level-1 fixed effect SE’s
    - Potential for contextual effects of level-1 predictors

Adding Predictors to a 2-Level Model for Persons in Groups

- Level-1 predictors (where time-varying predictors used to be) now refer to *Person-Level Variables*
  - Account for within-group (between-person) level-1 residual variance
- Level-2 predictors (where time-invariant predictors used to be) now refer to *Group-Level Variables*
  - Account for between-group level-2 random effects variances
- Same concerns about separation of between-group and within-group effects of level-1 predictors apply: NO SMUSHING ALLOWED
  - Grand-mean-centering or group-mean-centering of level-1 predictors (where group-mean-centering is like person-mean-centering from before)
  - Grand-mean-centering can be convenient for interpreting main effects, but gets more complicated when interpreting interaction terms…

Lecture 7:Two-Level Models for Clustered Data

- Modeling Dependency in Clustered Data
- **Grand-Mean-Centering vs. Group-Mean Centering in Multilevel Models for Clustered Data**
- Effect Size via Pseudo-R^2
- Clustered Data Example in SAS, SPSS, and STATA

3 Pieces of Information about Effects of Level-1 Predictors

- **Is the L2 Between-Group (BG) effect of X significant?**
  - Are groups higher on x_{ij} *on average* (GMx_j) higher on y_{ij} *on average*?
  - Only **GROUP**-mean-centering gives you this directly in the model
- **Is the L1 Within-Group (WG, Between-Person) effect of X significant?**
  - If you have higher x_{ij} values *than the rest of your group*, do you also have higher y_{ij} values than predicted *for the rest of your group*?
  - **Either GROUP- or GRAND-**MC with level-2 GMx_j gives you this directly
- **Are the BG and WG Effects of X of the same magnitude?**
  - The **L2 contextual effect**: is there *an additional bonus/decrement for each person's predicted* y_{ij} from being in a group that is high on x_{ij} on average (GMx_j) *above and beyond* (controlling for) the x_{ij} value for each person?
  - Only **GRAND**-mean-centering with level-2 GMx_j gives you this directly

** Can use **ESTIMATE** (in SAS) or **TEST** (in SPSS) or **LINCOM** (in STATA) or **NEW** (in Mplus) to get any implied effects not directly provided by the model parameters

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

Level 2		Original	Group-MC Level 1	Grand-MC Level 1
GMx_j	$GMx_j - 5$	x_{ij}	$WGx_{ij} = x_{ij} – GMx_j$	$L1x_{ij} = x_{ij} – 5$
3	-2	2	-1	-3
3	-2	4	1	-1
7	2	6	-1	1
7	2	8	1	3

Same GMx_j would go into the L2 model either way…

Under **Group-MC**, $\mathbf{WGx_{ij}}$ DOES NOT contain L2, BG variation, so GMx_j is uncorrelated with WGx_{ij}

Under **Grand-MC**, $\mathbf{L1x_{ij}}$ DOES contain L2, BG variation, so GMx_j IS correlated with $L1x_{ij}$

*** This means that the effects of $\mathbf{GMx_j}$ **and** $\mathbf{WGx_{ij}}$ **in Group-MC** will be the **same** regardless of whether the other effect is included, but that the effects of $\mathbf{GMx_j}$ **and** $\mathbf{L1x_{ij}}$ **in Grand-MC** will be **different** when they are together than their effects would be when included by themselves…

Clustered Data Model with **Group-MC for Level-1 X:** WG and BG Effects are Represented Directly

x_{ij} is **Group-MC** into WGx_{ij}, **WITH** GMx_j at Level 2:

L1: $y_{ij} = \beta_{0j} + \beta_{1j}(\mathbf{WGx_{ij}}) + \mathbf{e_{ij}}$

L2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(\mathbf{GMx_j}) + \mathbf{U_{0j}}$

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

WGx_{ij} contains only L1 variation (= x_{ij} – GMx_j)

GMx_j contains only L2 variation (= mean of x_{ij})

γ_{10} = WG effect of having higher X *than others in your group*

γ_{01} = BG effect of being from a group that has "high X" people

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

Clustered Data Model with **Grand-MC x_{ij} at Level 1 Only:** WG and BG Effects are **Smushed Together**

x_{ij} is **Grand-MC** into $L1x_{ij}$, **WITHOUT** GMx_j at L2:

L1: $y_{ij} = \beta_{0j} + \beta_{1j}(\mathbf{L1x_{ij}}) + \mathbf{e_{ij}}$

L2: $\beta_{0j} = \gamma_{00} + \mathbf{U_{0j}}$

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

$L1x_{ij}$ contains BOTH L1 and L2 variation

γ_{10} = <u>smushed*</u> WG and BG effect

***aka, *convergence*, *conflated*, *composite*, or *aggregate* effect**

So because $L1x_{ij}$ actually contains information about 2 different variables (WGx_{ij} and GMx_j), its 1 effect has to do the work of 2 predictors

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)

- Convergence effect will often be closer to the **within-group effect** (due to larger level-1 sample size and thus smaller SE), thus the problem with smushing is that the **between-group effect is wrong**
- **It is the rule, rather than the exception**, that between and within effects differ (Snijders & Bosker, p. 52-56, and personal experience!)
- However – you don't have to assume convergence in order to use grand-mean-centering for a level-1 predictor… here's how to fix it →

Clustered Data Model with **Grand-Mean-Centered Level-1 X:**

Tests Difference of WG and BG effects (It's been fixed!)

x_{ij} is **Grand-MC** into $L1x_{ij}$, **WITH** GMx_j at Level 2:

L1: $y_{ij} = \beta_{0j} + \beta_{1j}(L1x_{ij}) + e_{ij}$

L2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(GMx_j) + U_{0j}$

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

$L1x_{ij}$ contains BOTH L1 and L2 variation

GMx_j contains only L2 variation (=mean of L1)

γ_{10} *becomes* WG effect → pure L1 effect (now that GMx_j is included)

γ_{01} *is contextual (incremental)* L2 effect → tests difference of BG and WG effects

Because L2 variance is still in $L1x_{ij}$, GMx_j takes the <u>unique</u> part of the L2 effect that $L1x_{ij}$'s L2 variance didn't cover

Equivalence of Group- & Grand-MC (Fixed effects; Main effects only)

Group-Mean-Centering (uses WGx_{ij}):

L1: $y_{ij} = \beta_{0j} + \beta_{1j}(L1x_{ij} - GMx_j) + e_{ij}$

L2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(GMx_j) + U_{0j}$

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

$\rightarrow y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(L1x_{ij} - GMx_j) + U_{0j} + e_{ij}$ ← **In terms of WGx**

$\rightarrow y_{ij} = \gamma_{00} + (\gamma_{01} - \gamma_{10})(GMx_j) + \gamma_{10}(L1x_{ij}) + U_{0j} + e_{ij}$ ← **In terms of L1x**

Grand-Mean-Centering (uses $L1x_{ij}$):

L1: $y_{ij} = \beta_{0i} + \beta_{1j}(L1x_{ij}) + e_{ij}$

L2: $\beta_{0j} = \gamma_{00} + \gamma_{01}{}^{*}(GMx_j) + U_{0j}$

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

$\rightarrow y_{ij} = \gamma_{00} + \gamma_{01}{}^{*}(GMx_j) + \gamma_{10}(L1x_{ij}) + U_{0j} + e_{ij}$

Term	Group-MC	Grand-MC
Intercept	γ_{00}	γ_{00}
WG Effect	γ_{10}	γ_{10}
Contextual	$\gamma_{01} - \gamma_{10}$	$\gamma_{01}{}^{*}$
BG Effect	γ_{01}	$\gamma_{01}{}^{*} + \gamma_{10}$

Grand-Mean vs. Group-Mean Centering in Clustered Data

- Group- and Grand-MC are equal if level-1 effect is **fixed**
- Grand-MC may be convenient → L2 "contextual" effect

- Example: Effect of student SES on student achievement:

- Grand-Mean-Centering with school mean SES to avoid smushing:
  - **Level-1 WG Effect**: Effect of being rich kid relative to the rest of *your* school (after statistically controlling for school mean SES)
  - **Level-2 Contextual Effect**: **Extra** (incremental) effect of going to a rich school (**after statistically controlling for kid SES**)
- Group-Mean-Centering (with school mean SES by default):
  - **Level-1 WG Effect**: Effect of being rich kid relative to the rest of *your* school (already removed school mean SES from predictor)
  - **Level-2 BG Effect**: Effect of going to a rich school **NOT controlling for kid SES**

Parameter Interpretation across Methods of Centering Level-1 X

- **Group-mean-centering → subtract a VARIABLE**
  - L1 predictor $\mathbf{WGx_{ij}}$ = level-1 x_{ij} – original group mean x_j
    - *Directly represents Within-Group (WG) effect of X (regardless of whether GMx_j is in the model at L2)*
  - L2 predictor $\mathbf{GMx_j}$ = original group mean x_j – constant
    - *Directly represents Between-Group (BG) effect of X (regardless of whether WGx_{ij} is in the model, too)*

- **Grand-mean-centering → subtract a CONSTANT**
  - L1 predictor $\mathbf{L1x_{ij}}$ = original x_{ij} – constant
    - *WITHOUT GMx_j at L2, is combined BG and WG effects*
    - *WITH GMx_j at L2, becomes WG effect*
  - L2 predictor $\mathbf{GMx_j}$ = original group mean x_j – constant
    - *WITHOUT $L1x_{ij}$ at L1, is BG effect (like above)*
    - *WITH $L1x_{ij}$ at L1, becomes difference of BG and WG effects*

What about "Multilevel SEM"?

- In order to get BG and WG effects, so far we've separated the BG and WG variance in a level-1 predictor by brute force (e.g., by computing a GMx_j variable to use with $L1x_{ij}$ or WGx_{ij})
- An alternative is "multilevel SEM" (which isn't really SEM if it doesn't have other kinds of latent variables besides the MLM-based random effects, but whatever)
- Multivariate model → the variance in level-1 predictors is decomposed **by the model** into random intercept (BG) vs. residual (WG), the same as if it were an outcome (thus predictors = outcomes)
  - Pros:
    - Can have missing data on level-1 predictors (because are outcomes then)
    - Can be used to test multilevel mediation (currently impossible in MIXED)
    - May have less biased level-2 effects because there is no observed GMx_j variable assumed perfectly reliable (see Lüdke et al. 2008 Psych Methods)
  - Cons:
    - Greater estimation demands → more likely to blow up (only available in Mplus)
    - Different (but equivalent) syntax → BG or contextual effects (be careful)
    - Good luck fitting interaction terms! (→ latent variable interactions)

Interactions Involving Level-1 Predictors Belong at Both Levels…

Another example: Does the effect of employee motivation (x_{ij}) on employee performance interact with firm type (Z_j: profit, non-profit)?

- **Group-Mean-Centering for employee motivation (x_{ij}):**
  - WGx_{ij} by Z_j → Does the WG motivation effect differ by firm type?
  - GMx_j by Z_j → Does the BG motivation effect differ by firm type?
    - Moderation of total group motivation effect (not controlling for employee motivation)
    - If forgotten, then firm type moderates the motivation effect only at level 1 (pry **weird**)
- **Grand-Mean-Centering for employee motivation (x_{ij}):**
  - $L1x_{ij}$ by Z_j → Does the WG motivation effect differ by firm type?
  - GMx_j by Z_j → Does the contextual motivation effect differ by firm type?
    - 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 main effect of employee motivation has been un-smushed, the interaction of $L1x_{ij}$ by firm type would still be smushed, which assumes that firm type moderates the WG and BG motivation effects equally (pry **wrong**)

Interactions Involving Level-1 Predictors Belong at Both Levels…

Group-Mean-Centering (uses WGx_{ij}):

$$y_{ij} = \gamma_{00} + U_{0j} + e_{ij} + \gamma_{10}(L1x_{ij} - GMx_j) + \gamma_{01}(GMx_j) + \gamma_{02}(Z_j) + \gamma_{03}(GMx_j)(Z_j) + \gamma_{11}(L1x_{ij} - GMx_j)(Z_j)$$

←**In terms of WGx_{ij}**

$$y_{ij} = \gamma_{00} + U_{0j} + e_{ij} + \gamma_{10}(L1x_{ij}) + (\gamma_{01} - \gamma_{10})(GMx_j) + (\gamma_{02})(Z_j) + \gamma_{11}(L1x_{ij})(Z_j) + (\gamma_{03} - \gamma_{11})(GMx_j)(Z_j)$$

←**In terms of $L1x_{ij}$**

Grand-Mean-Centering (uses $L1x_{ij}$):

$$y_{ij} = \gamma_{00} + U_{0j} + e_{ij} + \gamma_{10}(L1x_{ij}) + \gamma_{01}^{*}(GMx_j) + \gamma_{02}(Z_j) + \gamma_{03}^{*}(GMx_j)(Z_j) + \gamma_{11}(L1x_{ij})(Z_j)$$

After adding an interaction with L2 Z_j 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*Z Effect:** $\gamma_{03} = \gamma_{03}^{*} + \gamma_{11}$ **Contextual*Z:** $\gamma_{03}^{*} = \gamma_{03} - \gamma_{11}$

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

Same is true for BG*WG intra-variable interactions (e.g., WGx_{ij}*GMx_j or $L1x_{ij}$*GMx_j)

Group-Mean-Centering (uses WGx_{ij}):

$$y_{ij} = \gamma_{00} + U_{0j} + e_{ij} + \gamma_{10}(L1x_{ij} - GMx_j) + \gamma_{01}(GMx_j) + \gamma_{02}(GMx_j)^2 + \gamma_{11}(L1x_{ij} - GMx_j)(GMx_j)$$

←In terms of WGx_{ij}

$$y_{ij} = \gamma_{00} + U_{0j} + e_{ij} + \gamma_{10}(L1x_{ij}) + (\gamma_{01} - \gamma_{10})(GMx_j) + (\gamma_{02} - \gamma_{11})(GMx_j)^2 + \gamma_{11}(L1x_{ij})(GMx_j)$$

←In terms of $L1x_{ij}$

Grand-Mean-Centering (uses $L1x_{ij}$):

$$y_{ij} = \gamma_{00} + U_{0j} + e_{ij} + \gamma_{10}(L1x_{ij}) + \gamma_{01}^{*}(GMx_j) + \gamma_{02}^{*}(GMx_j)^2 + \gamma_{11}(L1x_{ij})(GMx_j)$$

After adding an interaction with GM at both levels via GM^2, 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^2 Effect:** $\gamma_{02} = \gamma_{02}^{*} + \gamma_{11}$ **$Contextual^2$:** $\gamma_{02}^{*} = \gamma_{02} - \gamma_{11}$

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

Meaning of Random Effects in a 2-Level Model for Clustered Data

- In our previous examples, random effects were associated with time and other within-person level-1 predictors:
  - → e.g., Is the effect of time on health different across persons?
  - → e.g., Is the effect of stress on health different across persons?
- Similarly, in clustered data models, random effects can be associated with any level-1 (now person-level) predictor:
  - → e.g., Does the effect of child's SES on achievement differ across schools (i.e., does SES matter more for achievement in some schools)?
  - → Start with a fixed effect of X (then test if the X effect should be random)
- Remember: fixed and random effects mean different things!
  - → Fixed: Is there an effect that is different than 0, period?
  - → Systematically Varying: Does the effect vary systematically by predictors?
  - → Random: Does the effect differ randomly over groups?

When Group-MC vs. Grand-MC Matters:
Random Slopes of Level-1 Predictors

Group-Mean Centering:

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

$$y_{ij} = \gamma_{00} + \gamma_{01}(GMx_j) + \gamma_{10}(L1x_{ij} - GMx_j) + U_{0j} + \boxed{U_{1j}(L1x_{ij}) - U_{1j}(GMx_j)} + e_{ij}$$

Grand-Mean Centering:

$$y_{ij} = \gamma_{00} + \gamma^{*}_{01}(GMx_j) + \gamma_{10}(L1x_{ij}) + U_{0j} + \boxed{U_{1j}(L1x_{ij})} + e_{ij}$$

Both centerings yield equivalent models if the L1 effect is fixed, but NOT if the L1 effect is random.

The variance in GMx_j is NOT subtracted out of the random slope in Grand MC. Therefore, models with random slopes are not equivalent.

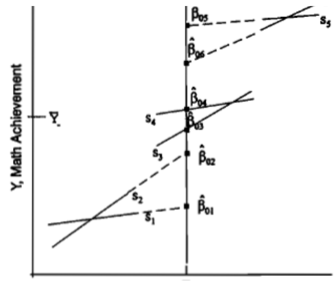
So which do you choose?

Adding Random Slopes of L1 X

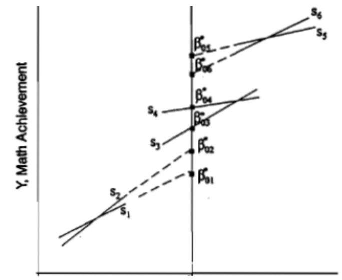
- **Random intercepts** mean different things under each model:
  - **Group-MC** → Individual differences at **WGx_{ij} =0** (every group should have)
  - **Grand-MC** → Individual differences at **$L1x_{ij}$=0** (not every group will have)
- **Differential shrinkage of the random intercepts** results 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 BG variation in X (more extrapolation needed)
  - Anecdotal example was presented in Raudenbush & Bryk (2002; chapter 5)
- Can also use ML AIC and BIC to decide which way to go…

Bias in Random Slope Variance

OLS Per-Group Estimates | **EB Shrunken Estimates**



Level-1 X



Level-1 X

Top right: Intercepts and slopes are homogenized in grand-MC

Bottom: Downward biased random slope variance in grand-MC

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

Lecture 7:Two-Level Models for Clustered Data

- Modeling Dependency in Clustered Data
- Grand-Mean-Centering vs. Group-Mean Centering in Multilevel Models for Clustered Data
- **Effect Size via Pseudo-R^2**
- **Clustered Data Example in SAS, SPSS, and STATA**

Variance Accounted For By Level-2 Predictors

- **Fixed effects of level 2 predictors *by themselves*:**
  - L2 (BG) main effects reduce L2 (BG) random intercept variance
  - L2 (BG) interactions also reduce L2 (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 L2 BG random slope variance
  - If the interacting level 1 predictor <u>not random</u>, any cross-level interaction with it will reduce the L1 WG residual variance instead
    - This is because the L2 BG random slope variance would have been created by decomposing the L1 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*:**
  - L1 (WG) main effects reduce L1 (WG) residual variance
  - L1 (WG) interactions also reduce L1 (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 L2 variance (e.g., Grand-MC predictors), then that L2 variance will also likely reduce L2 random intercept variance
  - If the level-1 predictor DOES NOT have L2 variance (e.g., Group-MC predictors), then its reduction in the L1 residual variance will cause an INCREASE in L2 random intercept variance
    - Same thing happens in Grand-MC, but you don't generally see it
    - It's just an artifact of estimation, and not a cause for alarm…
      here's a review of how it happens…

Negative Pseudo-R^2???

- As a consequence of reducing level-1 residual variance, the level-2 random intercept variance **(τ_{U0}^2) will usually INCREASE**
  – This is not really a problem, but here's how it happens:

- Observed **level-2 τ_{U0}^2** is NOT just between-group variance
  – Also has a small part of within variance (**level 1, σ_e^2**) due to sampling
    → **observed $\tau_{U0}^2 \approx$ true $\tau_{U0}^2 + (\sigma_e^2 / n)$**
    (as level-1 n goes up, effect of level-1 σ_e^2 is minimized)
  – ML estimate of "true" τ_{U0}^2 uses this as a correction factor:
    → **true $\tau_{U0}^2 \approx$ observed $\tau_{U0}^2 - (\sigma_e^2 / n)$**

- Example: observed $\tau_{U0}^2 = 4.65$, $\sigma_e^2 = 7.06$, $n = 4$
  – So **true τ_{U0}^2** = 4.65 – (7.06 / 4) = **2.88**
  – Add group-MC level-1 predictor, reduce σ_e^2 to 2.17
    - Pseudo-R^2 for σ_e^2 = (7.06 – 2.17)/7.06 = .69, or 69% reduction in σ_e^2
  – But τ_{U0}^2 was not reduced, so **true τ_{U0}^2** = 4.65 – (2.17 / 4) = **4.10** now

More on Pseudo-R^2 Effect Size

- Pseudo-R^2 is not calculated when adding **random** effects
  – Does not apply: fixed effects reduce variance, but random effects only **re-partition variance** (random effect = new pile of variance)
  – Calculate random effects confidence intervals instead!

- Pseudo-R^2 is only calculable across models with same piles of variance (meaning of each variance changes if others are added)
  – Another problem: Adding level-2 predictors of one random effect may cause other random effect variances to decrease through their correlation

- **A simple alternative: Total R^2**
  – Generate model-predicted y values from *fixed effects only* (using OUTPM in SAS, FIXPRED in SPSS, or PREDICT XB in STATA) and correlate with observed y values (then square that 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
    - Such as in cross-classified models… up next

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 (BG), L1 = Within-Group (WG; between-person)
  - L1 predictors are now person variables, and can have fixed, random, or systematically varying effects
  - L2 predictors are now group variables, and can have fixed or systematically varying effects
- Still no smushing allowed main effects of or interactions involving level-1 predictors:
  - Group-MC at L1: Get L1=WG and L2=BG effects directly
  - Grand-MC at L1: 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