# Stat 414 – Day 11
## Random Slopes cont., Centering (4.6, 5.2)

**Last Time:** Random slopes (Ch. 5)
- $Y_{ij} = \beta_{00} + u_{0j} + (\beta_{10} + u_{1j})x_{ij} + \epsilon_{ij}$
  - Effect of $x_{ij}$ differs among Level 2 groups (e.g., schools)
  - Adds variance component and covariance between slopes and intercepts
- Var($Y_{ij}$) = $\tau_0^2 + 2\tau_{01}x_{ij} + \tau_1^2 x_{ij}^2 + \sigma^2$
  - Results impacted by choice of scaling (origin) of *x* variables
  - Smallest when *x* = $-\tau_{01}/\tau_1^2$ (correlation vs. covariance)
  - Expect 95% of slopes within $\beta_{10} \pm 2\tau_1$
- Cov of two individuals in same group: $\tau_0^2 + \tau_{01}(x_{aj} + x_{bj}) + \tau_1^2(x_{aj}x_{bj})$
  - No simple ICC (depends on *x*)
- Explaining level 2 variance
  - Consider intercepts, slopes of "outcomes"

**Keep in mind:** With random slopes,

Corr($Y_{aj}$, $Y_{bj}$) = Cov($Y_{aj}$, $Y_{bj}$)/SD($Y_{aj}$) SD($Y_{bj}$)

$$= \frac{\tau_0^2+\tau_{01}(x_{aj}+x_{bj})+\tau_1^2(x_{aj}x_{bj})}{\sqrt{(\tau_0^2+2\tau_{01}x_{aj}+\tau_1^2x_{aj}^2+\sigma^2)(\tau_0^2+2\tau_{01}x_{bj}+\tau_1^2x_{bj}^2+\sigma^2)}}$$

For Netherlands data:

```
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 school   (Intercept) 0.28050  0.5296
          gevocab     0.01922  0.1386   -0.86
 Residual             3.66613  1.9147
```

```
getVarCov(model1b, type = "marginal")
cov2cor(cov$'767')
cov2cor(cov[[1]])
```

*Not a simple intraclass correlation coefficient with random slopes.*

**Example 1:** Day 10 Example 2 cont. (lmer)

```
Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 school   (Intercept) 3.6391814 1.90766
          gevocab     0.2605131 0.51040  -0.08
          age         0.0005558 0.02357  -0.69 -0.66
 Residual             3.6076583 1.89938

Fixed effects: geread ~ gevocab + age
                 Value Std.Error    DF
(Intercept)  2.9614055 0.4151887 10158
gevocab      0.5191496 0.0143563 10158
age         -0.0088390 0.0038396 10158
```

```
Random effects:
 Groups   Name        Variance   Std.Dev. Corr
 school   (Intercept) 0.10238213 0.319972
          cgevocab    0.01906890 0.138090  0.53
          cage        0.00002486 0.004986 -0.28 -0.96
 Residual             3.66396826 1.914149

Fixed effects:
             Estimate Std. Error t value
(Intercept)  4.343884   0.032704 132.825
cgevocab     0.519277   0.014361  36.159
cage        -0.008882   0.003822  -2.324
```

**Definition: *Grand mean centering*** is subtracting off the overall mean, $y_{ij} - \bar{y}$. Benefits include interpretation of intercepts, reduce collinearity in "product terms." Can also be especially helpful with strong slope/intercept covariance. "When working with random coefficient models, it is best that ***all*** independent variables be centered.

***Group mean centering*** is subtracting off the group mean, $y_{ij} - \bar{y}_j$. This is not equivalent and impacts interpretation, but is less common in practice.

(a) Can we include an interaction between age and gevocab in the two cases? How do we interpret this interaction?

(b) How does centering *gevocab* impact our earlier analysis?

(c) What if don't want the random slopes to be correlated?

**Example 2:** Recall the data predicting *PctBush* with counties clustered within states. We want to consider the percentage of families in the county in a traditional nuclear form as predictor.

(a) What is the overall mean *PctFamily*?

(b) What are the means of *PctFamily* for the first four states?

AL AR AZ CA

(c) What would PctFamily be if grand-mean centered? Group-mean centered?

| state_po | county | PctBush | Income | PctFamily | Grand mean centered | Group mean centered |
|---|---|---|---|---|---|---|
| AL | Autauga | 75.67352 | 48.458 | 25.3 | | |
| AR | Boone | 66.27191 | 34.974 | 19.6 | | |
| CA | Orange | 59.80009 | 64.611 | 26.1 | | |

(d) In a simple model with a fixed slope for PctFamily, how do the slope and intercept compare depending on which version you use? Why?

| | Intercept | Slope |
|---|---|---|
| Raw data | | |
| Grand mean | | |
| Group mean | | |

(e) In a model with a random slope for PctFamily, how do the slope and intercept and slope/intercept correlation values compare for the different versions? What about the prediction for Autauga?

| | Intercept | Slope | Corr | $\hat{y}$ |
|---|---|---|---|---|
| Raw data | | | | |
| Grand mean | | | | |
| Group mean | | | | |

(f) Which models are equivalent in the predicted values? Why?

In considering Level 2 variables to add to the model, an “especially important type” of “contextual variable” is the group mean.

(g) Include pctfamilyavg in the model, along with the interaction with PctFam.

(h) Include StatePctFam, grand-mean centered, in the model and compare interpretations depending on how the Level 1 PctFam variable is centered.

| | Intercept | PctFam | StatePctFam | |
|---|---|---|---|---|
| Raw data | 17.4 | 0.89 | 1.22 | |
| Both Grand mean | 57.5 | 0.89 | 1.22 | |
| L1: Group mean/<br>L2: Grand mean | 57.5 | 0.89* | 2.11 | |

Note: Testing the group level coefficient (coefficient of the group mean) is referred to as the Hausman specification test in econometrics.

(i) What does a positive coefficient for the group mean tell you?

(j) What do you notice about the sum of the slope coefficients in row 2 and the slope coefficients in row 3?

(k) How do these results compare to the “within group” regression and “between group” regression?

(l) How do you summarize and evaluate the significance of the “contextual effect”?

**Example 3:** Reconsider the Netherlands data on language score based on verbal IQ from the text (e.g., Section 4.6)

(a) Fit the model with both IQ and mean IQ. What is the slope of the “within group” regression?

(b) What is the additional contribution to langPOST from the class mean verbal IQ? Is this statistically significant?

(c) What if we put both grand-mean centered IQ and grand-mean centered mean IQ in the model; how do we interpret the coefficient of mean IQ in this model?

(d) How does the coefficient of mean IQ change when we use the group-mean centered verbal IQ?

(e) How does this coefficient compare to the slope of mean IQ in the aggregated data set?

(f) Interpret the slope/intercept correlation in the multilevel model.

(g) What about including SES in the model?