# Stat 414 – Day 9
## Three-level Random Intercepts Model (4.9)

**Last Time:**
- In a random intercept model, the Level 2-specific intercepts are modeled as having a mean and variance, which in turn can be used to generate the Best Linear Unbiased Prediction (BLUP) for each Level 2 unit if you are interested in the expected outcome for each group.
  - The BLUP-based predicted outcomes are sometimes called "shrinkage" estimates because they are "shrunk" toward the mean intercept for Level 2 units with few observations.
- These estimates are sometimes sorted to estimate "value added" by each unit, but with some controversy (http://www.amstat.org/asa/files/pdfs/POL-ASAVAM-Statement.pdf)

**Example 1:** Data were collected to predict reading achievement for 10,903 third-grade students nested within 568 classrooms nested within 160 schools (achieve.txt). Let's fit a simple "null model" of random intercepts (also called "unconditional means" model).

(a) Write out the multilevel model and a "diagram" of the data structure.

(b) Record and interpret the four parameter estimates.

(c) Does the school a student attends appear to have an impact on their reading scores? Does the classroom?

(d) What is the total variance in student scores? What percentage of the total variation is at each level of the model? (VPCs)

The general form of the intraclass correlation coefficient:

ICC = (sum of variance components pertinent to two observations) / (sum of all variance components including error variance)

(f) What is the intraclass correlation coefficient for two students in the same class?

(g) What is the intraclass correlation coefficient for two students in the same school?

(h) What is the estimated correlation of two students at different schools?

(i) What is the correlation of two classes in the same school? Are classes in the same school more similar than classes at different schools?

(j) Do we need the random class effect if we have adjusted for the school-to-school differences? Compare the model with classes nested in schools to just the model with school random effects. What do you learn? What do you recommend?

(k) What about a fourth level! Is there significant variation at the *corp* (district) level? How many school corporations are there?

**Example 2:** Data (in the DAAG package, science dataset) were collected on 1385 Australian students from 20 classes in 12 private schools and 46 classes in 29 public schools. The response variable is "like," a measure of how much the students like Science (scores range 1-12). Compare the provided models/commands in the R script. Which match and which do not? Why not? Write a summary of each model and which are correct.

**Example 3: Adding Fixed Effects**

(a) For the achieve data, add gevocab (general vocabulary score) to the two-level model. Is this a Level 1 or Level 2 variable? Is it statistically significant?

(b) After accounting for *gevocab*, how did the estimate of the school-to-school variation in intercepts change? Within-school variation?

**Pseudo-$R^2$ values**

$R^2$ =

In multilevel setting, there isn't a great analog, but several has been proposed, to use cautiously. ("the literature does not seem to have converged on this topic.")

| | | | |
|---|---|---|---|
| Level 1 | $1 - \hat{\sigma}^2(\text{full}) / \hat{\sigma}^2(\text{null})$ | $1 - \dfrac{(\hat{\sigma}^2(\text{full}) + \hat{\tau}^2(\text{full}))}{(\hat{\sigma}^2(\text{null}) + \hat{\tau}^2(\text{null}))}$ | 1 – L(full)/L(null) |
| Level 2 | $1 - \hat{\tau}^2(\text{full}) / \hat{\tau}^2(\text{null})$ | $1 - \dfrac{(\hat{\sigma}^2(\text{full})/n + \hat{\tau}^2(\text{full}))}{(\hat{\sigma}^2(\text{null})/n + \hat{\tau}^2(\text{null}))}$ | |
| | Could be negative | | |

(c) How much variation in reading scores does *gevocab* account for above and beyond the null model?

Return to the three-level model and add student vocabulary score (gevocab), number of students in class (clenroll), and number of students in school (cenroll).

(d) Is this a better fitting model? Are any of these variables significant? How did the variance components change? Are they still significant? What does that tell you? How do we interpret them?

(e) Calculate an "$R^2$" for this model by comparing it to the null (three-level) model.

(f) Does including an interaction between vocabulary score and school size improve the fit of the model? How would you interpret this interaction?

(g) What would it mean for there to be an interaction between vocabulary score and school?

**Example:** (from J. Miles, RAND Corporation)
Brooks et al. (2008) studied incentives to improve attendance in adult literacy classes. Classes in the UK were assigned to either a 5 £ voucher for each class attended or a 20 £ voucher for taking the final exam. One response variable was number of class sessions attended.

(a) Why do you think the study was set up this way? What are possible consequences?

(b) Is the treatment group statistically significant? (indicator coding)

```
> summary(lm(sessions ~ 1 + group, data = adultlit))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.2805     0.3419  15.446  < 2e-16 ***
group         1.4052     0.5038   2.789  0.00597 **
```

(c) What is the intraclass correlation coefficient?

```
> anova(lm(sessions ~ 1 + classid))
Analysis of Variance Table

Response: adultlit$sessions
                   Df Sum Sq Mean Sq
adultlit$classid   27 743.33 27.5307
Residuals         124 768.87  6.2006
```

In terms of Mean Squares

ICC: $\dfrac{\text{MSgroup} - \text{MSError}}{\text{MSgroup} + (n\text{-}1)\ \text{MSError}}$

where $n$ is the average group size

```
> lmer(sessions ~ 1 + (1 | classid)
Random effects:
 Groups   Name        Std.Dev.
 classid  (Intercept) 2.029
 Residual             2.491
Number of obs: 152,
groups:  classid, 28
```

(d) What is the effective sample size?

(e) How does this change our standard error for the group effect?