Stat 414 – Day 7
Estimating Group Effects (4.8)

Last Time:
- Multilevel data
  - Aggregation (level 2 analysis) vs. Disaggregation (level 1 analysis)
- Multilevel models
  - Induced correlation, Intraclass correlation coefficient
  - Level 1 and Level 2 equations vs. Composite equation
  - lmer and lme commands in R
- In the empty ("null") multilevel model, our parameters were β_0, σ^2, τ^2. The random effects aren't considered parameters but "unobservable latent effects."

If we treat the Level 2 grouping variable as random, how are the Level 2 effects estimated? Assumption: $u_j \sim N(0, \tau^2)$.

Example 1: Suppose the bball.txt file is the batting averages for 6 players over several seasons.
(a) What is the overall mean batting average for these 43 observations?

(b) What are the means and standard deviations for each player? What is a rough 95% confidence interval for Jones' "population" mean batting average?

(c) Do you really think Rodriguez and Suarez are that much better than everyone else?

(d) Treat *player* as a fixed effect (with effect coding) and fit a linear model. What are the estimated overall mean batting average across all players? Is this the same as in (a)? Why not? Where does it come from?

(e) What does this model estimate for the "effect" of Jones? For Jones' population mean batting average? Is this the same as in (b)? New 95% confidence interval?

(f) Now treat *player* as random. What is the intraclass correlation coefficient? The estimate of the overall mean?

Suppose we were still interested in estimating the individual players' "effects"/expected outcomes for each Level 2 group.

(g) Find the estimated batting averages for Jones and Suarez. How do the estimates compare to those from model1? Which one changed more? Why do you think that is?

We can think of each random effect estimate (predicted group mean) as being a weighted average of the overall mean and the group mean: *w*(group mean) + (1- *w*)(overall mean) where the weights depend on the relative sizes of the variance components and on the group sizes, $w_j = \tau^2 / (\tau^2 + \sigma^2/n_j)$. The impact of this is "partial pooling."

(f) Verify the estimated group means for Suarez and Jones

Suarez weight: Jones weight:

Suarez estimate: Jones estimate:

(g) What happens to the weight as the group size increases?

(h) What does "no pooling" mean?

(i) What does "complete pooling" mean?

(j) What if σ^2 gets smaller?

(k) What if τ^2 gets smaller?

(l) Explain how "no pooling" overestimates the player to player variation.

Example 2: (Example 4.1 in text): We want to predict language test scores (langPOST) in Grade 8 students (~ age 11) in elementary schools in the Netherlands.

(a) Identify the fixed and random effects in this context. Identify level 1 and level 2.

(b) Use R to create the null model. How many students and how many schools are in the dataset?

(c) How much variation in the language scores is there from school to school? A statistically significant amount? What is the intraclass correlation coefficient?

(d) What do you predict for the language score of a randomly selected student? Is this the same as the mean of all the language scores in the sample? Why or why not?

(e) What standard error would you put around your estimate? Is this the same as the standard deviation of all the language scores in the sample? Why or why not?

(f) Is it reasonable to pick out the schools with the largest positive residuals and conclude they are doing something better than the other schools?

Example 3: The following are modeling results for a randomized controlled trial at 29 clinical centers. The dependent variable is diastolic blood pressure.

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Intercept	centre	10.6671
Residual		73.7126

.

Solution for Fixed Effects						
Effect	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		90.9656	0.7836	28	116.09	<.0001
treat	A	3.1134	0.6382	1061	4.88	<.0001
treat	B	1.4145	0.6439	1061	2.20	0.0283
treat	C	0	.	.	.	.

(a) What is an estimate of the ICC? Calculate and interpret

(b) What is the patient level variance?

(c) What is the center level variance?

(d) What is an estimate of the ICC? Calculate and interpret.

(e) What is the expected diastolic blood pressure for a randomly selected patient receiving treatment C at a center with average aggregate blood pressure scores?

(f) What is the expected diastolic blood pressure for a randomly selected patient receiving treatment C at a center with aggregate blood pressure scores at the 16th percentile?

(g) What is the expected diastolic blood pressure for a randomly selected patient receiving treatment B at a center with aggregate blood pressure scores at the 97.5th percentile?

(h) What is the expected diastolic blood pressure for a randomly selected patient receiving treatment A at a center with aggregated blood pressure scores at the median?