Stat 414 - Day 7
Random Intercepts Models (Ch. 4)

Previously:
- Fixed vs. Random Effects: If categories themselves aren't so much of interest, but want to consider the "grouping units" as a random sample from a larger population, can treat as random effects. Also helpful if group sizes are small, can "borrow information" across groups to estimate individual effects.
- $E(Y_{ij}) = \beta_0 + u_j + \epsilon_{ij}$ where we are assuming $\epsilon_{ij} \sim N(0, \sigma^2)$, including $cov(\epsilon_{ij}, \epsilon_{kj}) = 0$, and $u_j \sim N(0, \tau^2)$ and $cov(\epsilon_{ij}, u_j) = 0$
- Benefits include fewer parameters to estimate and generalizability to larger population of units. Also *models* the correlation of observations within groups.
- Results in "partial pooling"

Example 1: Reconsider our swim stroke data with fixed effects (cap/no cap, 4 swim strokes, 4 swimmers, effect coding).

Model 1: $Y_i = \beta_0 + \beta_1 swimmer_{1i} + \beta_2 swimmer_{2i} + \beta_3 swimmer_{3i} + \epsilon_i$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  18.2528     0.5017  36.413  < 2e-16
ID1          -2.3153     0.8682  -2.667  0.01258
ID2          -2.8278     0.8682  -3.257  0.00295
ID3           0.1459     0.8682   0.168  0.86773
```

```
Analysis of Variance Table

Response: Time.sec.
          Df Sum Sq Mean Sq F value    Pr(>F)
ID         3 306.80 102.268  12.718 1.995e-05
Residuals 28 225.15   8.041
```

Residual SE = 2.836; R^2 = 0.577; ICC: (102.268-8.041)/(102.268+7*8.041)=0.594

(a) Which is larger, the between group variation or the within group variation?

(b) Let y_{ij} represent the time of the i^{th} swim of the j^{th} swimmer. Rewrite the statistical model using swimmer as random effect. Include model assumptions about the error terms.

$Y_{ij} = \beta_0 + u_j + \varepsilon_{ij}$ $\varepsilon_{ij} \sim N(0, \sigma^2)$ composite
$u_j \sim N(0, \tau^2)$

In "multilevel language," we will refer to the swims as the Level 1 units and the swimmers as the Level 2 units. Alternatively, we can write the model as:

swims Level 1: $Y_{ij} = \beta_{oj} + \epsilon_{ij}$ with $\epsilon_{ij} \sim N(0, \sigma^2)$
swimmers Level 2: $\beta_{oj} = \beta_{00} + u_j$ with $u_j \sim N(0, \tau^2)$

Because there are no explanatory variables in this model, we will refer to it as the "null model." This allows us to first partition the variation into within vs. between groups and is generally the starting point to which we will compare future models.

```
Linear mixed-effects model fit
  Data: swimdata
       AIC     BIC    logLik
  169.6901 173.992 -81.84503

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev:    3.431958 2.835658
```

```
Fixed effects:  Time.sec. ~ 1
               Value Std.Error DF  t-value
(Intercept) 18.25281  1.787698 28 10.21023
```

$\hat{\beta}_0$ sample to sample variability in $\hat{\beta}_0$'s

(c) How many parameters are estimated by this model? 3

$\hat{\tau}$ $\hat{\sigma}$ ← within swimmer

τ between swimmer from pop'n mean swimmer

(d) What is the estimated intercept? How do we interpret it? What is the standard error of the intercept? How has it changed?

$\hat{\beta}_0 = 18.25 = \bar{y}$ estimate of avg swim time

$SE(\hat{\beta}_0) = 1.79$ increased when RE model

(e) What is the estimated "within group" variation? What is the estimated "between group" variation? According to this model, what fraction of the total variation is due to the swimmer-to-swimmer variation?

within $= \hat{\sigma}^2 = 2.836^2$ between $= 3.4319^2$ $\frac{3.4319^2}{2.836^2 + 3.4319^2} = .594$

To decide whether you have significant swimmer-to-swimmer variation, you can

- Use the original fixed-effects ANOVA anova(model1)
- Use a LRT test (using gls) to compare the models with and without the swimmers
- Examine confidence intervals for τ intervals(model1R)

(f) What do you learn from the confidence interval output?

Did not include zero $\rightarrow \tau > 0$

The confidence intervals are a little more "controversial" and different packages may approach these methods a little differently. All of them are aiming to test $H_0: \tau^2 = 0 \; vs. H_a: \tau^2 > 0$. The fact the variance can never be zero can occasionally lead to "boundary conditions" but usually something you don't have to worry about. You could also cut the p-value in half to reflect the one-sided alternative.

(g) What is the difference between the "marginal" variance-covariance matrix from the model and the "conditional" variance-covariance matrix?

19.869 marginal $= \hat{\sigma}^2 + \hat{\tau}^2$ total variation

8.041 conditional $= \hat{\sigma}^2$ within group

Computer Problem 7: Data were collected from nine beaches along the Dutch coast. Five readings were taken for each beach, measuring the species richness (number of different species).

(a) Fit the null model with both lme and lmer.

(b) Discuss the similarities and differences between the two outputs. Do they use ML or REML?

(c) What is the ICC?

(d) What is the correlation of two observations on the same beach? What is the correlation of two observations from different beaches?

(e*) Explain what each standard error calculation represents/what information is/is not used in each.

Example 2: Netherlands language scores (see text) The Netherlands Language dataset examines language test scores (langPOST) in Grade 8 students (~ age 11) for elementary schools in the Netherlands. (See p. 50 for more information about this dataset.) Students (Level 1) are nested within a random sample of schools (Level 2) and we will treat the schools as random effects. Create the null model (using lmer for graph below)

```
library(lme4)
nullmodel = lmer(langPOST ~ 1 + (1|schoolnr), data = neth, REML = FALSE)
#using ML to better match the output in the text
performance::icc(nullmodel)
```

(a) Based on the above output, how many students are in the data set? How many schools are in the dataset?

3758 211

(b) What do you learn from the ICC? Which is larger the within group or between group variation?

22.4% of variation is between schools

(c) Using the null model, what do you predict for the language score of a randomly selected student? Is this the same as the mean language score in the dataset? Why or why not?

$\hat{\beta}_0 = 41.00 \neq \bar{y} = 41.41$, unequal group sizes

(d) What is the estimated standard deviation of the language scores? Is this the same as the standard deviation of all the language scores in the sample? Why or why not?

$\sqrt{\hat{\sigma}^2 + \hat{\tau}^2} = 8.72 \neq s_y = 8.89$

Consider the first estimated random effect

```
ranef(nullmodel)$schoolnr[1,] ## [1] -4.043693
```

(e) How do you interpret this value?

School 1's avg score is below school average

Consider the distribution of estimated random effects

```
hist(ranef(nullmodel)$schoolnr[,])
plot(ranef(nullmodel))
```

(f) Do the school effects appear to follow a roughly normal distribution?

yes, roughly

(g) (Based on the output) What is the estimated standard deviation of this normal distribution? (Check the histogram to make sure you answer is reasonable)

4.257

4.275 (population)
+ school effects ⇒

(h) What do you predict for the average language score for a school in the 84th percentile? (Hint: What is special about the 84th percentile in a normal distribution?)

84% of schools have on "effect" below 4.257

$41 + 4.257 = 45.257$

predicted $\bar{y}_j$ for 84%

A "Catepillar plot" is a nice visual for sorting and visualizing the estimated effects.
(i) Is it reasonable to pick out the schools with the largest positive effects and conclude they are doing something better than the other schools? (Hint: This is more of an opinion question, check out http://www.amstat.org/asa/files/pdfs/POL-ASAVAM-Statement.pdf to learn about some of the controversy surrounding Value Added Models)

not "causal" so be cautious

Notes:

- Treating person as a fixed effect would "fail to reflect uncertainty resulting from variation among people." That's why the standard errors tend to be smaller. With random effects we are able to make inferences about the population of swimmers, not just these four, a more difficult task.
- Random effects are also helpful when the group sizes are small they allow more