Stat 414 - Day 16 (Ch. 4!)
Random Intercepts

Last Time:

- 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 $cov(\epsilon_{ij}, \epsilon_{ik}) = 0$, and $u_j \sim N(0, \tau^2)$ and $cov(\epsilon_{ij}, u_j) = 0$
- Benefits include fewer parameters to estimate and generalizabiity to larger population of units. Also *models* the correlation of observations within groups.

Example 1: Brooks et al. (2008) studied incentives to improve adult literacy. Twenty-eight classes were assigned to either receive the treatment group (participants received a 5£ (US $10) M&S voucher for each class they attended) or to a control group. The main outcome of interest was number of class sessions attended.

(a) Identify the response variable of interest.

(b) Identify some potential sources of variation in this response variable.

```
model1 <- lm(sessions ~ group, data = adultlit); summary(model1)
```

(c) What is the estimated treatment effect? (0 = intervention, 1 = control), standard error, and *t*-statistic?

(d) What model assumption is violated in this analysis? Why do you think the data were collected that way? What is the potential consequence?

This model assumes that all of the observations are independent, including students in the same class. Is there much class-to-class variability? In other words, how correlated are the observations in each class?

```
icc =
```

(e) Find the "effective sample size" and use that value to "correct" the standard error of the coefficient. Does it impact the statistical significance of that coefficient in this case?

```
#Effective sample size: N/(1 + (n-1)×ICC)  fixing the typo in Day 13
ess =

#'corrected' standard error
#sqrt(n/ess) * SE(beta-hat)
```

To account for this dependence, we can instead carry out a "random effects anova": $y_{ij} = \beta_0 + u_j + \epsilon_{ij}$ for the i^{th} observation in the j^{th} group. Let's rewrite this as a multilevel model:

Level 1: $y_{ij} = \beta_{0j} + \epsilon_{ij}$ with $\epsilon_{ij} \sim N(0, \sigma^2)$

Level 2: $\beta_{0j} = \beta_{00} + u_j$ and $u_j \sim N(0, \tau^2)$ and $cov(\epsilon_{ij}, u_j) = 0$

Notice that substituting β_{0j} back into the Level 1 equation returns us to the original "composite" equation. This is also sometimes referred to as a "random intercepts" model or a "null model" or a "variance components model." Use this model to recalculate the ICC.

```
model0 <- lme(fixed = sessions ~ 1, random = ~ 1 | classid, data = adultlit)
summary(model0)
```

(e) What are the estimated coefficients? The ICC?

We can view the estimated variance-covariance matrix for individual classes (see Rmd file).

Is the variability between the classes statistically significant?

```
#unfortunately this doesn't work
#MLmodel <- lme(fixed = sessions ~ `, data = adultlit)
#but you can go back to gls (REML)
GLSmodel0 <- gls(sessions ~ 1 , data = adultlit)
anova(model0, GLSmodel0)
#what does this tell you?
intervals(model0)
```

Although, even if the variability is not statistically significant, accounting for classes is definitely the more "valid" model. So now let's consider the treatment effect, after adjusting for class.

```
MLmodel1 <- lme(sessions ~ group, random =  ~1 | classid, data = adultlit)
summary(MLmodel1)
```

(f) How does adjusting for this lack of independence impact the standard error of the slope coefficient from model 1?

Notes

- If the observations within a group are correlated, then we are violating the "independence" condition and possibly artificially reducing our standard errors (makes it look like we have a lot more observations than we really do).
- Multilevel models "find the right standard error." We also *model* the dependence rather than just "correcting" for it.

To think about:

We will actually make more use of

```
library(lme4)
MLmodel1b = lmer(sessions ~ group + (1 | classidF), data = adultlit)
```

What's the same/different in the output?

```
#install.packages("ggeffects")
library(ggeffects)
library(tidyverse)

library(lme4)
MLmodel1b = lmer(sessions ~ group + (1 | classidF), data = adultlit)

ggpredict(MLmodel1b, terms = c("group", "classid [sample = 9]"), type = "re") %>%
plot()
```