Stat 414 - Day 6
Random vs. Fixed Effects

Last Time:

- Observations within groups/clusters/subjects are often correlated with each other.
- The intraclass correlation coefficient is one way to measure that correlation. Similar to R^2 it sees how much of the total variation is between groups/clusters/subjects (how reliably can we identify which group an observation is from) and can be interpreted as the correlation between two randomly selected observations from the same group/cluster/subject.
- If you have a larger intraclass correlation coefficient, the effective sample size is smaller.

Example 1: Caffeine cont. When we looked at just the participants variable, the ICC was 0.7877. We also note $\bar{y}$ = 474 (taps/min?) and $SD(Y)$ = 24.647 taps/min ($V(Y)$ = 607.45). & group means 497, 459, 470, 510 = $\hat{y}_i$ from model B
So we are now changing the assumptions of the basic regression model.

$$cor(\epsilon_{ij}) = \rho \neq 0$$

So what does our variance-covariance matrix look like? First we need to remember the difference between covariance and correlation.

$$Cor(X,Y) = Cov(X,Y)/[SD(X)SD(Y)]$$

(a) What are we considering to be X and Y?

Pairs of observations from each person

(b) What are our estimates for $SD(X)$ and $SD(Y)$?

Could use $\hat{\sigma}$ for both

(c) What is the estimated covariance?

Cov = Cor x SD x SD = $\hat{\rho}\hat{\sigma}^2$

Approach 1: So we can go to generalized least squares rather than OLS to incorporate this assumption into our model. The following model assumes "compound symmetry" (equal variances, equal covariances).

```
modelC <- nlme::gls(Taps ~ 1 , corr = corCompSymm(form = ~1 | participant))
nlraa::var_cov(modelC)
```

(d) What is the estimated intercept? Why? Its standard error? How many parameters are estimated? What is $\hat{\sigma}$?

$\hat{\beta}_0 = 474 = \bar{y}$ w/ $SE(\hat{\beta}_0) = 19.23$
3 parameters: β_0, σ, & ρ $\hat{\sigma} = 26.63$ ($\approx SD(Y)$)

(e) Summarize what you learn from the variance-covariance matrix (pattern). What do the diagonal elements represent? The off-diagonal? Which entries are zero and why?

$\hat{\sigma}^2 = 709$ (diagonal values)
Cov = 709/(26.63)(26.63) = 558.7 when same person,
zero if different people

Let's use a likelihood ratio test to compare these two models

```
modelD <- gls(Taps ~ 1, data=fingertapstudy)
nlraa::var_cov(modelD)
anova(modelC, modelD)
```

$y_i = \beta_0 + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$
2 parameters: β_0 & σ^2
$\hat{\sigma}^2 = 607.56 = Var(Y)$
note 607.56 < 709

$H_0: \rho = 0$ $H_a: \rho \neq 0$
df = 1 (ρ) small p-value
⇒ evidence against H_0

(f) Where does 607.56 come from? What is the null hypothesis of this test? What are the df of this test? What do you conclude?

The intraclass correlation is statistically significant.

So then we included participant in the model and found a statistically significant subject-adjusted association between stimulant type and finger tapping rate. But this study is a good example where we aren't really all that interested in the four participants themselves, we were just trying to control for that person-to-person variability, to help us assess the person-adjusted differences among the stimulants. In fact, we might be willing to consider the participants as a random sample?…

(g) Suppose we had a larger study with lots more participants. What would be a downside to including the participant variable in the model?

having to use a whole bunch of indicator variables (df)

Random Effects: In a situation like this, one option is to treat person as a *random effect* rather than a *fixed effect*. This means we are going to treat these 4 participants not as (the only) 4 levels of a factor, but as a random sample from a population (if I did the study again, I would get 4 different participants). The assumption we are going to make is that the "participant effects" follow a normal distribution, centered at zero, with variance τ^2. Let's call these participant effects, u_j, so we have $u_j \sim N(0, \tau^2)$. Our model equation becomes: $Y_{ij} = \beta_0 + u_j + \epsilon_{ij}$ where $u_j \sim N(0, \tau^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$. We also assume $cor(u_j, \epsilon_{ij}) = 0$. ($j^{th}$ observation on i^{th} subject)

u_i

Big deal, I changed β's to u's, but that is one way of saying we aren't considering the participant effects as parameters anymore. Instead, we replace them with one parameter, τ^2, which represents the participant-to-participant variation in the population of (potential) participants. This "small" change will have a large impact on the properties of the model.

(h) According to this model, what is $V(Y_{ij})$? $= V(\beta_0 + u_j + \epsilon_{ij}) = V(u_j + \epsilon_{ij}) = \tau^2 + \sigma^2$

(i) According to this model, what is $Cov(Y_{ij}, Y_{ik})$? $= cov(u_i + \epsilon_{ij}, u_i + \epsilon_{ik})$

$= cov(u_i, u_i) + cov(\epsilon_{ij}, \epsilon_{ik}) = \tau^2 + 0 = \tau^2$

To fit this model, today we will use the "lme" command from the nlme package

```
model4 = lme(fixed = Taps ~ 1 , random = ~1 | participant, data = fingertapstudy, method="REML") #The notation (1|subject) is how we tell R to treat the participants as random effects
```

(h) How many parameters are estimated in this model? How does the estimated intercept change? Standard error? What are the estimated variance components?

3: $\beta_0 + \sigma^2 + \tau^2$ (rather than 3 β's for the 4 subjects)

$\hat{\beta}_0 = 474$ w/ SE = 12.53 (like d), $\hat{\sigma} = 12.67$, $\hat{\tau} = 23.43$

We can view the estimated variance-covariance matrix for individual subjects.

```
getVarCov(model4, subject = "1", type = "marginal")[[1]]
```

And we can make R do the conversion to correlations

```
cov2cor(getVarCov(model4, subject = "1", type = "marginal")[[1]])
```

So we have partitioned the total random variability into a variance component for the individual observations within each person (assumed to be the same across the participants) and a variance component for the participants. This also nicely induces a non-zero correlation between two observations from the same Level 2 units (this allows us to model dependence within the groups).

(i) Find the estimated "total variation" by summing $\hat{\tau}^2 + \hat{\sigma}^2$. $12.27^2 + 23.63^2 = 709$

(j) How much of this variation is due to the different participants?

$23.63^2 / 709 = .788$

Shrinkage Estimation: Even though we say we are not all that interested in the individual u_j and that they aren't really parameters but "unobservable latent effects," we do still get estimates for them that are used to estimate τ and it might still be interesting to explore those estimates (e.g., do they appear to be normally distributed?) But how are they estimated differently?

```
#The output no longer gives us the estimated effects for the players, but R does
store them for us.
ranef(model4)        -15.59, -13.76, -3.67, 33.03
#Fitted values, prediction for each player
fits=predict(model4); fits      498.4, 460.2, 470.3, 507.0
```

(k) Compare the estimated means from this model to model B (fixed effects)

Each is a bit closer to $\bar{y}$

Definitions: One way to estimate a participant's effect is to ignore all the other participant, call this *no pooling*. Another way is to ignore the player to player differences and use the overall mean, call this *complete pooling*. Treating the player as a random effect creates *partial pooling*. We can think of each predicted group mean as being a weighted average of the group mean and the overall mean: $w(groupmean) + (1 - w)(overallmean)$ where the weight for group $j(w_j)$, depends on the relative sizes of the variance components and on the group size, $w_j = \tau^2/(\tau^2 + \sigma^2/n_j)$. The weights reflect the "reliability" of the group.

(l) Calculate the weight for participant 1? How will it compare to participant 2?

$23.63^2/(23.63^2 + 12.67^2/3) = .912$, same for each participant

(m) Summarize what you learn from the following graph. Which mean appears to change the most and which the least? Why is that?

You can think of the red dots as being 91.2% of the way towards the group mean. So the red and blue dots will look pretty similar (not much shrinkage because the person to person variation was so large in relation to the within person variation), but person D looks so different just because their effect was so large (the group mean so far from the overall mean)

Example 2: Suppose we have batting averages for 6 players over several seasons.

	Missing <chr>	n <chr>	Mean <chr>	SD <chr>
Anderson	0	6	0.295	0.032
Jones	0	11	0.202	0.029
Mitchell	0	6	0.323	0.075
Rodriguez	0	6	0.550	0.045
Smith	0	11	0.357	0.042
Suarez	0	3	0.550	0.050

observed sample means & SDs

$\hat{\beta}_0 = .3796 = (.295 + \cdots + .550)/6$ rather than $\bar{y}$ (.344)

(a) Do you expect a high or low ICC value? Explain.

Fairly high because again repeat observations on same individual, but now across seasons rather than immediately so not as high as last example?

(b) Opinion: Do you really think Rodriguez and Suarez are that much better than everyone else? What else could be going on? Which of these averages do you find the most/least "trustworthy"? Why?

could be a fluke, esp w/ 3 seasons but might believe Jones & Smith more with 11 seasons

(c) How do these player estimated effects compare to model 1 (fixed effects)?

Predicted means: .296, .204, .324, .347, .357, .544

All pulled closer to the intercept

(d) Whose estimated effect (Jones or Suarez) changed more? Why does that make sense for these data?

Suarez (.544 vs .547) because of smaller sample size

(e) Calculate the weights for Jones and Suarez. Why are the weights pretty large? Which is larger? Why?

Jones: $.140^2/(.140^2 + .0448^2/11) = .991$

Suarez: $.140^2/(.140^2 + .0448^2/3) = .967$ Both are fairly large but Jones has less shrinkage

(f) Verify the estimated group means for Jones and Suarez using these weights. Which changes (from the group mean) more? Why?

Jones = $.991^2 \times .202 + .009 \times .3796 = .204$

Suarez = $.967^2 \times .550 + .033 \times .3796 = .544$

Notes:

- Some packages/functions report the estimated variances, some the estimated standard deviations, some both.
- This gives us yet another interpretation of ICC as the percentage of variation in the response at Level 2 rather than Level 1.
- Degree of shrinkage depends on the variance of the effect (τ) and the number of observations per level in the effect. With large variance estimates, there is little shrinkage.
    - You can consider fixed effects as a special case of random effects where the variance component is very large.
- If variance component is small, then more shrinkage.
    - If the variance component is zero, the group effects are shrunk to exactly zero. It is even possible to obtain highly negative variance components where the shrinkage is reversed.
- If very few observations per level, then more shrinkage.
- If many observations per level, the estimates shrink less.
    - You can consider fixed effects as a special case with infinitely many observations.