Stat 414 - Day 12
Multiple Random Slopes (4.6, 5.2)

Last Time: Modeling Heteroscedasticity

- Adding random slopes induces unequal diagonal values and non-identical off-diagonal in the variance-covariance matrix (marginal vs. conditional)
- $Var(Y_{ij}) = \tau_0^2 + 2\tau_{01}x_{ij} + \tau_1^2 x_{ij}^2 + \sigma^2$
  - Variance of response is quadratic function of explanatory variable
  - Impacted by choice of scaling (origin) of x variable
  - Smallest when $x = -\tau_{01}/\tau_1^2$ (correlation vs. covariance)
- Cov of two individuals in the same group: $\tau_0^2 + \tau_{01}(x_{ij} + x_{kj}) + \tau_1^2(x_{ij}x_{kj}))$
  - Depends on the corresponding x-values
  - No simple ICC (depends on x)
  - Simplest: use x = 0
- $Corr(Y_{ij}, Y_{kj}) = \frac{Cov(Y_{ij}, Y_{kj})}{SD(Y_{ij})SD(Y_{kj})} = \frac{\tau_0^2 + \tau_{01}(x_{aj} + x_{bj}) + \tau_1^2(x_{aj}x_{bj})}{\sqrt{(\tau_0^2 + 2\tau_{01}x_{aj} + \tau_1^2 x_{aj}^2 + \sigma^2)(\tau_0^2 + 2\tau_{01}x_{bj} + \tau_1^2 x_{bj}^2 + \sigma^2)}}$

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).
A couple of data cleaning issues:
• I'm pretty sure 1=male and 2= female (you might want to convert to 0/1)
• The classes are not uniquely numbered but the schools are

(a) What is the largest source of variation in these students' reading scores?

students w/in schools $\hat{\sigma}^2$

(b) Interpret the correlation between the slopes and intercepts (Hints: Use the graph. Make sure you could do this in context).

$\hat{\tau}_{01} = -.86 \Rightarrow$ larger intercepts tend to have smaller slopes

(c) What if we center gevocab first?

```
model2 <- lmer(geread ~ gevocab + age + (1 + gevocab + age| class), data = achieve)
## Random effects:
##  Groups   Name        Variance     Std.Dev.  Corr
##  class    (Intercept) 0.0029100432 0.0539448
##           gevocab     0.0000005703 0.0007552 -0.96
##           age         0.0000012455 0.0011160 -0.91  0.87
##  Residual             3.8595254633 1.9645675
## Number of obs: 10320, groups:  class, 8
```

(c) Can we add age to the model? With random slopes? How many parameters does this add to the model? Is age significantly related to reading scores? If so, how? How do the two slope variances compare? How do you interpret the correlations?

While I think adding age worked ok, even with random slope, this particular dataset does sometimes get "unstable" pretty easily. Adding age with random slopes and 4 parameters (fixed effect, variance of slopes, correlation with intercepts and other slope effects). The slope variances are small (so probably not needed and adds to convergence issues) but is larger for age than gevocab) *(d) What if we didn't want the random slopes to be correlated?*

```
#First try
model6 <- lmer(geread ~ cgevocab + age + (cgevocab|school)  + (age | school), data
= achieve)
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  school   (Intercept) 0.0328289 0.18119
##           cgevocab    0.0196512 0.14018   1.00
##  school.1 (Intercept) 2.6720526 1.63464
##           age         0.0002237 0.01496  -0.99
##  Residual             3.6601374 1.91315
## Number of obs: 10320, groups:  school, 160
##
## Fixed effects:
```

a little awkward because now have two sets of random intercepts

```
#What about
summary(lmer(geread ~ cgevocab + age + (1|school) + (-1 + cgevocab|school) + (-1 +
age|school), data=achieve), corr=F)
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  school   (Intercept) 0.09566  0.3093
##  school.1 cgevocab    0.01834  0.1354
##  school.2 age         0.00000  0.0000
##  Residual             3.66710  1.9150
## Number of obs: 10320, groups:  school, 160
##
```

one set of random intercepts and no correlation among the level 2 random effects. Not that now the variance of the slopes of age is zero!

```
Computing profile confidence intervals ...
                2.5 %    97.5 %
.sig01      0.4544115 0.5983430
.sig02      0.1675981 0.4083006
.sig03      0.3147424 0.5430904
.sigma      2.1710534 2.2328418
(Intercept) 4.1838496 4.4679558
```

This output is for the 4-level model proposed next...

(e) Do we need all of the levels?

Notes:

- The number of correlation terms is equal to the number of unique pairs among Level Two random effects
- With three of more levels, will distinguish between *variance partitioning coefficients (vpc)* and *intraclass correlation coefficient (ICC)* = (sum of group variances for the individual)/(total variation)
    - Two students in same class ($\sigma_u^2 + \sigma_v^2$) vs. two students at same school but different classes (σ_v^2)