Stat 414 - Day 17
Three-level models (4.9)

Last Time: Using AR(1) structure at Level 1

- With longitudinal data, may want to also consider alternative correlation structures for the Level 1 residuals
- AR(1) assumes $\epsilon_{ij} = \rho\epsilon_{(i-1)j} + v_{ij}$ which models $cor(\epsilon_{ij}, \epsilon_{(i-1)j}) = \rho$ but still assumes equal variance σ^2.
- $var(y_{ij}) = \tau_0^2 + \sigma^2$
- $cov(y_{ij}, y_{kj}) = \tau_0^2 + \rho^{|i-k|}\sigma^2$
- Three-level models

Example 1: Case Study (from Finch, Bolin, & Kelly)

Data were collected to predict reading achievement for 10,903 third-grade students nested within 568 classrooms nested within 160 schools (achieve.txt).

```
achieve = read.table("https://www.rossmanchance.com/stat414F20/data/Achieve.txt" ,
header=TRUE)
# I have reason to believe gender = 1 is female and gender = 2 is male
```

Unconditional means model

Fit the unconditional means three-level (null) model, putting the higher level first to see how much variation is at each level.

```
#library(lme4)
summary(model0 <- lmer(geread~ (1|school/class), data = achieve, REML = F))
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: geread ~ (1 | school/class)
   Data: achieve

      AIC       BIC    logLik -2*log(L)  df.resid
    46150     46179    -23071     46142     10316

Scaled residuals:
   Min     1Q Median     3Q    Max
-2.305 -0.629 -0.210  0.304  3.867

Random effects:
 Groups       Name        Variance Std.Dev.
 class:school (Intercept) 0.273    0.522
 school       (Intercept) 0.309    0.556
 Residual                 4.847    2.202
Number of obs: 10320, groups:  class:school, 568; school, 160

Fixed effects:
            Estimate Std. Error t value
(Intercept)   4.3081     0.0548    78.6
```

```
#model0b <- lme(geread ~ 1, random = ~1 | school/class, data = achieve, method="ML"
)
#summary(model0b)
```

(a) How many parameters are in this model? How do you interpret them?

4: overall intercept, sigma, class intercept variance, school intercept variance

What if we try

```
summary(lmer(geread~ (1|school) + (1 |class), data = achieve))
Linear mixed model fit by REML ['lmerMod']
Formula: geread ~ (1 | school) + (1 | class)
   Data: achieve

REML criterion at convergence: 46267

Scaled residuals:
   Min     1Q Median     3Q    Max
-2.294 -0.636 -0.214  0.285  3.886

Random effects:
 Groups   Name        Variance Std.Dev.
 school   (Intercept) 0.3906   0.6250
 class    (Intercept) 0.0052   0.0721
 Residual             5.0424   2.2455
Number of obs: 10320, groups:  school, 160; class, 8

Fixed effects:
            Estimate Std. Error t value
(Intercept)   4.3301     0.0642    67.5
```

(b) What is the issue? How can we “fix” this?

ⓘ Code

```
achieve$school_class <- paste(achieve$school,"_", achieve$class)
summary(model1 <- lmer(geread~ (1|school) + (1 |school_class), data = achieve))
Linear mixed model fit by REML ['lmerMod']
Formula: geread ~ (1 | school) + (1 | school_class)
   Data: achieve

REML criterion at convergence: 46146

Scaled residuals:
   Min     1Q Median     3Q    Max
-2.305 -0.629 -0.209  0.305  3.867

Random effects:
 Groups       Name        Variance Std.Dev.
 school_class (Intercept) 0.273    0.522
 school       (Intercept) 0.312    0.558
 Residual                 4.847    2.202
```

```
Number of obs: 10320, groups:  school_class, 568; school, 160

Fixed effects:
            Estimate Std. Error t value
(Intercept)    4.308      0.055    78.3
```

(c) What percentage of the total variation is at each level? (VPC = variance partitioning coefficient)

ⓘ Code

```
#don't be fooled by the name!
performance::icc(model1, by_group=TRUE)
# ICC by Group

Group        |   ICC
--------------------
school_class | 0.050
school       | 0.057
#The largest source of variation in the scores is among students in the same clas
s.
```

(d) What do you learn from the following?

```
model1 = lmer(geread~ (1|corp/school/class), data = achieve, REML = F)
summary(model1)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: geread ~ (1 | corp/school/class)
   Data: achieve

      AIC       BIC    logLik -2*log(L)  df.resid
    46110     46146    -23050     46100     10315

Scaled residuals:
   Min     1Q Median     3Q    Max
-2.299 -0.631 -0.213  0.303  3.944

Random effects:
 Groups            Name        Variance Std.Dev.
 class:school:corp (Intercept) 0.2754   0.525
 school:corp       (Intercept) 0.0869   0.295
 corp              (Intercept) 0.1734   0.416
 Residual                      4.8470   2.202
Number of obs: 10320, groups:
class:school:corp, 568; school:corp, 160; corp, 59

Fixed effects:
            Estimate Std. Error t value
(Intercept)   4.3258     0.0715    60.5
anova(model1, model0)
Data: achieve
Models:
model0: geread ~ (1 | school/class)
```

```
model1: geread ~ (1 | corp/school/class)
       npar   AIC   BIC logLik -2*log(L) Chisq Df Pr(>Chisq)
model0    4 46150 46179 -23071     46142
model1    5 46110 46146 -23050     46100  42.2  1    8.1e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#confint(model1)
```

Variation in intercepts across corp (think school district) also appears to be meaningful.

(e) What do you learn from the following? (Better fitting model? Any of the variables significant? Variance explained? Do signs of coefficients make sense? (gevocab = student vocabulary scores, clenroll = size of the student's reading class, cenrol = size of the student's school)

```
summary(model2 <- lmer(geread ~ gevocab + clenroll + cenroll + (1 | school/class),
data = achieve, REML = F), corr=FALSE )
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: geread ~ gevocab + clenroll + cenroll + (1 | school/class)
   Data: achieve

      AIC       BIC    logLik -2*log(L)  df.resid
    43101     43152    -21544     43087     10313

Scaled residuals:
   Min     1Q Median     3Q    Max
-3.220 -0.568 -0.208  0.319  4.473

Random effects:
 Groups       Name        Variance Std.Dev.
 class:school (Intercept) 0.0902   0.300
 school       (Intercept) 0.0739   0.272
 Residual                 3.6975   1.923
Number of obs: 10320, groups:  class:school, 568; school, 160

Fixed effects:
               Estimate  Std. Error t value
(Intercept)  1.67617679  0.20694549    8.10
gevocab      0.50773055  0.00842293   60.28
clenroll     0.01890214  0.00950530    1.99
cenroll     -0.00000371  0.00000361   -1.03
fit warnings:
Some predictor variables are on very different scales: consider rescaling
anova(model2, model0)
Data: achieve
Models:
model0: geread ~ (1 | school/class)
model2: geread ~ gevocab + clenroll + cenroll + (1 | school/class)
       npar   AIC   BIC logLik -2*log(L) Chisq Df Pr(>Chisq)
model0    4 46150 46179 -23071     46142
model2    7 43101 43152 -21544     43087  3055  3     <2e-16 ***
```

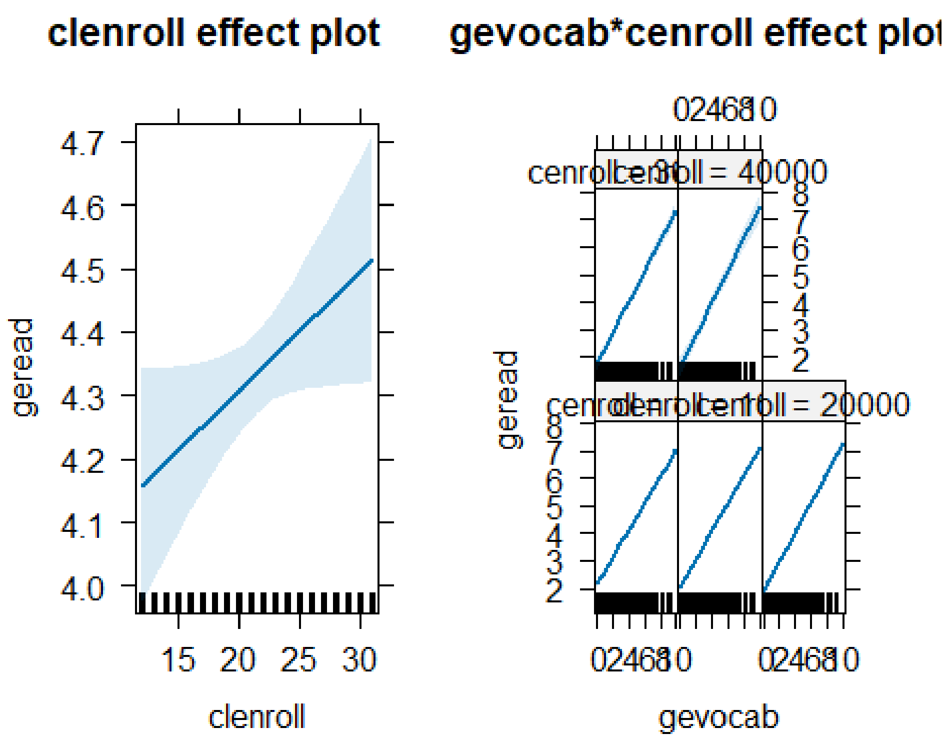
```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#confint(model2)

#performance::r2(model2, by_group= TRUE)
m2 <- lmer(geread ~ gevocab + clenroll + cenroll + (1|school) + (1|school_class), d
ata = achieve)
performance::r2(m2, by_group= TRUE)
# Explained Variance by Level

Level        |    R2
--------------------
Level 1      | 0.237
school       | 0.755
school_class | 0.668
```

The AIC has decreased (46150 to 43101) and the improvement in the model fit appears to be statistically significant (p-value from LRT < .0001). The student vocabulary scores (Level 1) and the students' reading class (level 2) appear to be significant, after adjusting for the other variables, but the size of the school size does not (level 3). The negative coefficient on the size of the class may seem counter intuitive, but larger classes may also come with additional TA support. The model (compared to the null model) explains 23.7% of the variation at Level 1 (within class), (1 - .09016/0.2726) => 66.8% at the class level and 75.5% at the school level compared to the null model. (Note, we had to use the split up version to get these values for each level)

(f) What do you learn from the following?

```
summary(model3 <- lmer(geread ~ gevocab + clenroll + cenroll +  gevocab:cenroll + (
1 | school/class), data = achieve), corr=F)
Linear mixed model fit by REML ['lmerMod']
Formula: geread ~ gevocab + clenroll + cenroll + gevocab:cenroll + (1 |
    school/class)
   Data: achieve

REML criterion at convergence: 43152

Scaled residuals:
   Min     1Q Median     3Q    Max
-3.190 -0.568 -0.206  0.318  4.472

Random effects:
 Groups       Name        Variance Std.Dev.
 class:school (Intercept) 0.0886   0.298
 school       (Intercept) 0.0751   0.274
 Residual                 3.6982   1.923
Number of obs: 10320, groups:  class:school, 568; school, 160

Fixed effects:
                   Estimate  Std. Error t value
(Intercept)      1.75154296  0.20999286    8.34
```

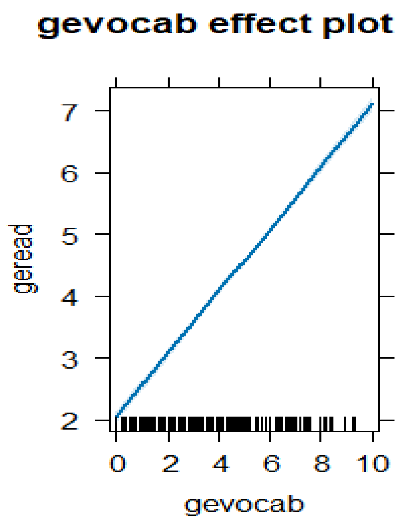
```
gevocab          0.48999982  0.01168332   41.94
clenroll         0.01880075  0.00951172    1.98
cenroll         -0.00001315  0.00000563   -2.34
gevocab:cenroll  0.00000234  0.00000107    2.19
fit warnings:
Some predictor variables are on very different scales: consider rescaling
anova(model2, model3)
Data: achieve
Models:
model2: geread ~ gevocab + clenroll + cenroll + (1 | school/class)
model3: geread ~ gevocab + clenroll + cenroll + gevocab:cenroll + (1 | school/class
)
       npar   AIC   BIC logLik -2*log(L) Chisq Df Pr(>Chisq)
model2    7 43101 43152 -21544     43087
model3    8 43099 43157 -21541     43083  4.81  1      0.028 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
library(ggeffects)
plot(effects::allEffects(model3))
```

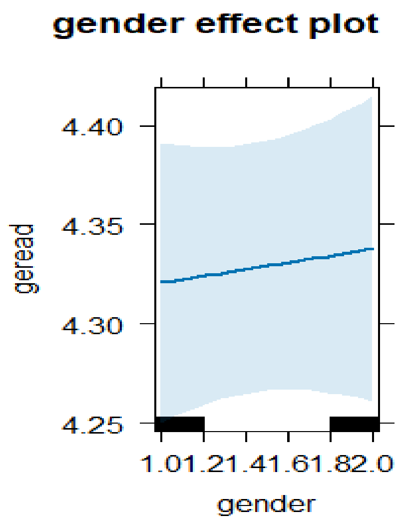


The interaction between student vocabulary score and school enrollment is statistically significant (LRT p-value = .02829) and the coefficient is positive: The rate of increase in z_read with each one-unit increase in gevocab is larger for larger schools than for smaller schools, after adjusting for the other variables. Or ‘the impact of gevocab on reading scores varies depending on the size of the school.’ We note that the impact of school size (assuming gevocab = 0) is now significant.

(g) What about?

```
summary(model4 <- lmer(geread ~ gevocab + gender + (1 + gender| school/class), data
= achieve))
Linear mixed model fit by REML ['lmerMod']
Formula: geread ~ gevocab + gender + (1 + gender | school/class)
   Data: achieve

REML criterion at convergence: 43108

Scaled residuals:
   Min     1Q Median     3Q    Max
-3.204 -0.568 -0.207  0.317  4.445

Random effects:
 Groups       Name        Variance Std.Dev. Corr
 class:school (Intercept) 0.1483   0.3851
              gender      0.0196   0.1399   -0.62
 school       (Intercept) 0.0328   0.1810
              gender      0.0066   0.0812   0.61
 Residual                 3.6925   1.9216
Number of obs: 10320, groups:  class:school, 568; school, 160

Fixed effects:
            Estimate Std. Error t value
(Intercept)  2.01559    0.07560   26.66
gevocab      0.50909    0.00841   60.55
gender       0.01723    0.03922    0.44

Correlation of Fixed Effects:
        (Intr) gevocb
gevocab -0.527
gender  -0.757  0.039
optimizer (nloptwrap) convergence code: 0 (OK)
Model failed to converge with max|grad| = 0.0093274 (tol = 0.002, component 1)
plot(effects::allEffects(model4))
```



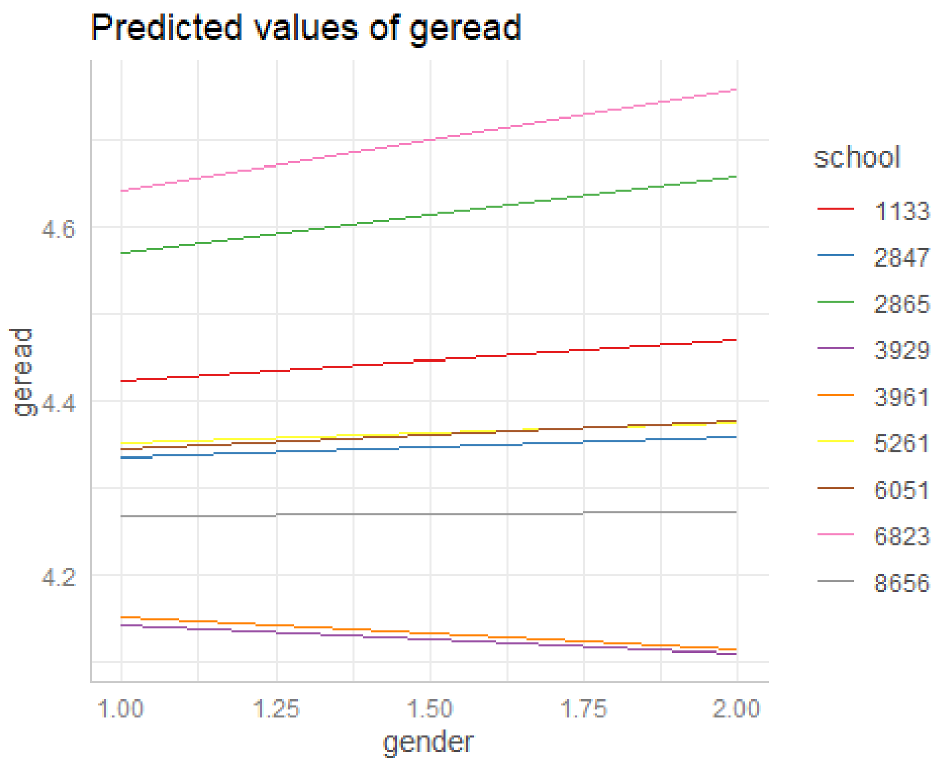



```
#Note in these data, the males are predicted (after averaging over everything) to
have higher reading scores than the females (with same vocab score and no school or
class random effects)
```

There is meaningful variation in the intercepts (average scores when gender = 0 with average gevocab) at both the class level and the school level (though less so). But there is not a significant gender gap, after adjusting for gevocab (t-value = 0.439), and there is not a lot of variation in the slopes of gender among classes or among schools.

From the graphs (now below), I think I can interpret the correlations of the random effects. At the school level, the correlation between the intercepts and slopes is positive. So school that have higher intercepts (gender = 0) tend to have steeper slopes (larger benefits for males compared to females).

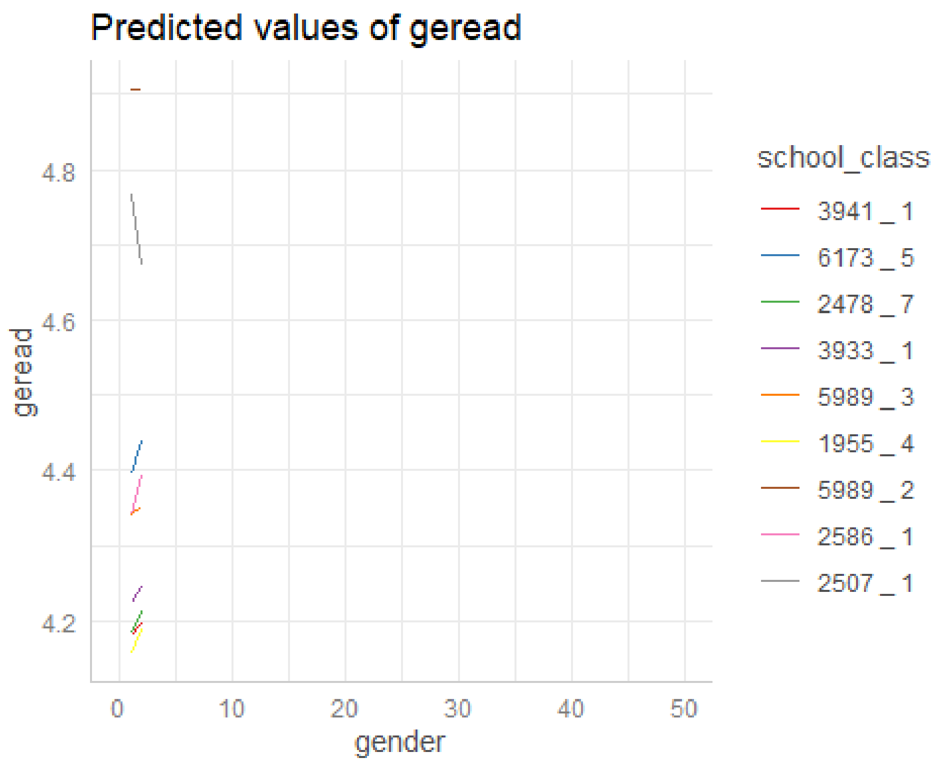
```
plot(ggpredict(model4, terms = c("gender", "school [sample = 9]"), type = "random")
, show_ci = FALSE)
```



For the classes, the correlation between the random intercepts and random slopes is negative, with a minimum at -(-0.57)/.01349 = 42, so the lines are (in general) fanning in and the classes with larger intercepts at gender = 0 tend to have smaller differences between males and females (perhaps even reversing direction). So schools with higher means for females tend to have lower scores for males and classes with lower means for females tend to have higher scores for males. You can also think of this as less variation in the class means for males than for females.

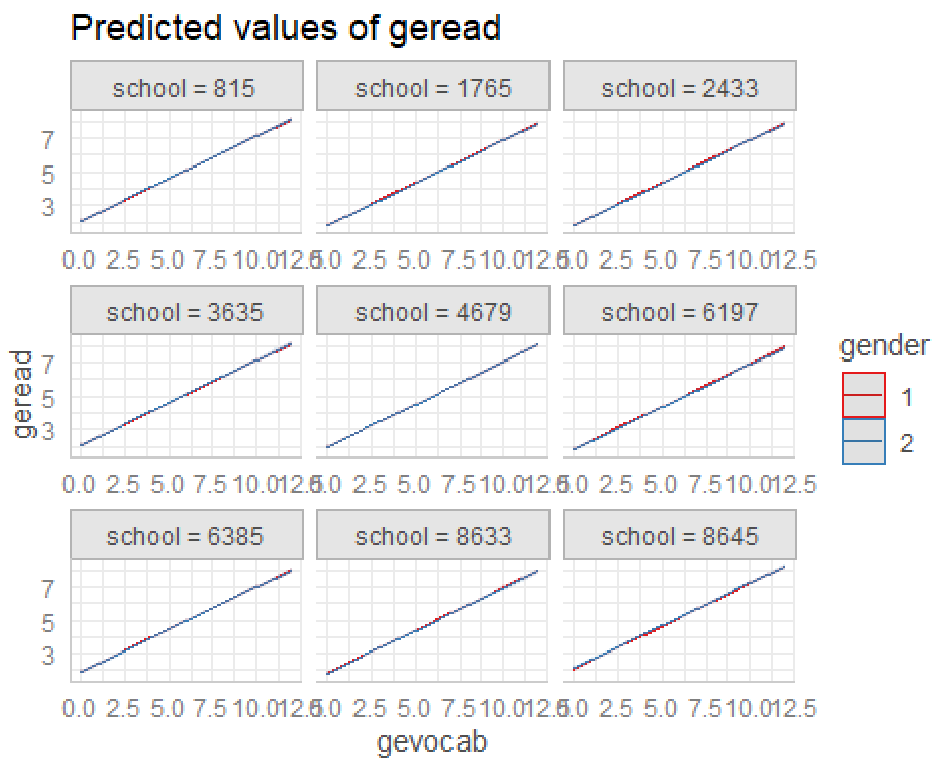
But in a practical sense, these lines are pretty much parallel here!

```
#refitting the model with the 'separate' error components
summary(model4 <- lmer(geread ~ gevocab + gender + (1 + gender|school) + (1 + gende
r | school_class), data = achieve))
library(ggplot2) #to make the ridiculous scaling
plot(ggpredict(model4, terms = c("gender", "school_class [sample = 9]"), type = "ra
ndom"), show_ci = FALSE) +
  scale_x_continuous(limits = c(0, 50))
```

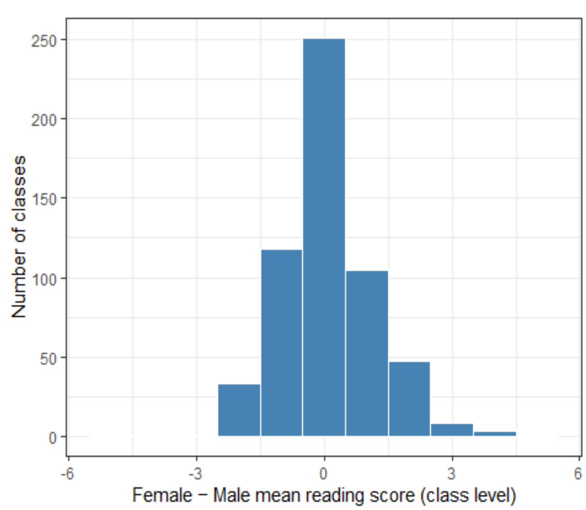


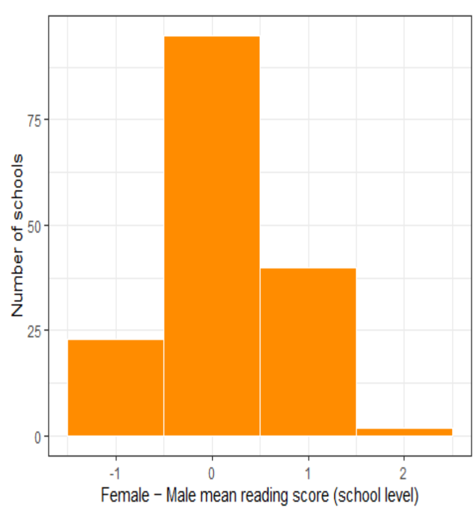
Here is another interesting graph, reinforcing how minute the gender difference is!

```
plot(ggpredict(model4,
          terms = c("gevocab", "gender", "school [sample = 9]"),
          type = "random"))
```



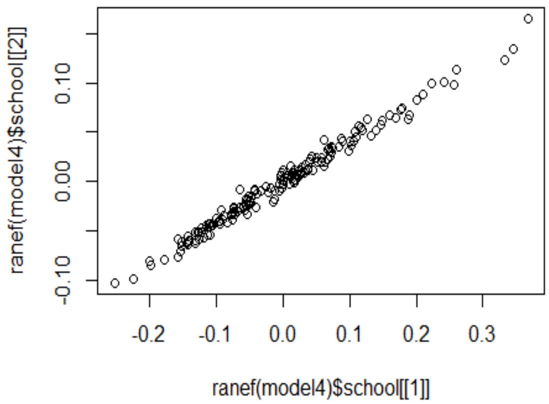
Here is a comparison of the "gender gaps" across classes and schools. So the main thing I notice is that we are almost as likely to have postive differences as negative differences at the class level, but generally more positive differences (male - female) at the school level. So this is telling us that at the class level, it's the classes with higher female averages that tend to have those negative differences.



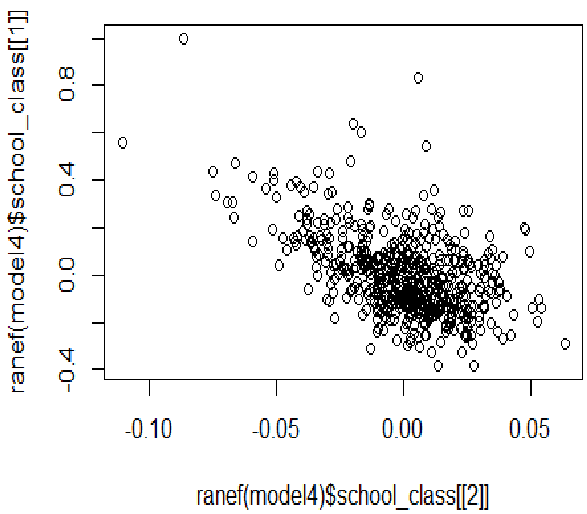


And the graphs of random slope effects vs. random intercept effects!

```
plot(ranef(model4)$school[[2]]~ranef(model4)$school[[1]])
```




```
plot(ranef(model4)$school_class[[2]], ranef(model4)$school_class[[1]])
```



(h) How would you interpret the following models?

```
lmer(geread~gevocab+gender + (1|school) + (gender|class), data = achieve)
Linear mixed model fit by REML ['lmerMod']
Formula: geread ~ gevocab + gender + (1 | school) + (gender | class)
   Data: achieve
REML criterion at convergence: 43140
Random effects:
 Groups   Name        Std.Dev. Corr
 school   (Intercept) 0.3145
 class    (Intercept) 0.0507
          gender      0.0015   1.00
 Residual             1.9404
Number of obs: 10320, groups:  school, 160; class, 8
Fixed Effects:
(Intercept)      gevocab       gender
     1.9996       0.5132       0.0198
optimizer (nloptwrap) convergence code: 0 (OK) ; 0 optimizer warnings; 1 lme4 warni
ngs
```

```
lmer(geread~gevocab+gender + (-1 + gender|school) + (1|class), data = achieve)
Linear mixed model fit by REML ['lmerMod']
Formula: geread ~ gevocab + gender + (-1 + gender | school) + (1 | class)
   Data: achieve
REML criterion at convergence: 43146
Random effects:
 Groups   Name        Std.Dev.
 school   gender      0.1926
 class    (Intercept) 0.0567
 Residual             1.9415
Number of obs: 10320, groups:  school, 160; class, 8
Fixed Effects:
(Intercept)      gevocab       gender
     2.0019       0.5150       0.0148
lmer(geread~gevocab+gender + (1|corp) + (1|school) + (gender|class), data = achieve
)
Linear mixed model fit by REML ['lmerMod']
Formula: geread ~ gevocab + gender + (1 | corp) + (1 | school) + (gender |
    class)
   Data: achieve
REML criterion at convergence: 43106
Random effects:
 Groups   Name        Std.Dev. Corr
 school   (Intercept) 0.17074
 corp     (Intercept) 0.25160
 class    (Intercept) 0.04478
          gender      0.00496  1.00
 Residual             1.94044
Number of obs: 10320, groups:  school, 160; corp, 59; class, 8
Fixed Effects:
(Intercept)      gevocab       gender
      2.017        0.511        0.019
```

Question 1: This gives us random intercepts at the school level and random intercepts and random slopes for gender at the class level. The level 2 and level 3 residuals are assumed to be uncorrelated, but the random intercepts and random slopes at level 2 are assumed correlated. So we have 8 parameters (intercept, slope of gevocab, slope of gender, tau for school intercepts, tau for gender intercepts, tau for gender slopes, covariance between those, and sigma. (You get a bump up or down depending on your class and depending on your school)
Question 2: This fits random intercepts at the class level and allows gender effects to vary across schools but does not fit additional random intercepts at the school level. (You get a bump up or down depending on your class but not again depending on your school.) Our parameters will be sigma, overall intercept, slope for gevocab, slope for gender, tau for gender, and tau for class (6 parameters). No correlations (only one random effect at each level). Question 3: This has random intercepts at all 3 levels and random slopes for gender at the class level. So the parameters are sigma, intercept, slope for gevocab, slope for gender, tau for corp, tau for school, tau for class, tau for gender-class and covariance between class intercepts and class slopes = 9 parameters. This would be equivalent to geread ~ gevocab + gender + (1 | corp/school/class) + (-1 + gender | class)

This command lmer(geread~gevocab+gender + (1 + gender |corp/school/class) , data=achieve)) means we also have random slopes for gender at the corp and school levels as well (and correlations between those slopes and the intercepts at each level; 9 + 4 = 13 parameters)

(i) How would you interpret the following models? How many parameters in each model?

lmer(z_read ~ cen_pos + female + cen_size + (1 | region/schoolid))

schools nested within regions; random intercepts for schools and random intercepts for regions; parameters: intercept, 3 fixed effects, variance for school intercept, variance for region intercept, sigma = 7

lmer(z_read ~ cen_pos + female + cen_size + (1 | region) + (1 | schoolid)

same as first as long as school ids are unique

lmer(z_read ~ cen_pos + female + cen_size + (1 + female | region/schoolid)

ntercept, 3 fixed fixed effects, random intercepts for school and random intercepts for region, random slopes for female for school and region, 2 covariance, sigma: 11

lmer(z_read ~ cen_pos + female + cen_size + (1 + female | schoolid) + (1 | region)

intercept, 3 fixed fixed effects, random intercepts for school and random intercepts for region, random slopes for female for school , 1 covariance, sigma: 9

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 shared group variances for the individuals)/(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)