

# Homework 2

*Leah Schultz*

*9/28/2017*

## Chapter 3: Growth Curves

```
library(dplyr)
library(tidyr)
library(lme4)
library(ggplot2)
library(broom)
library(merTools)
purpose <- read.csv("~/Dropbox/Lab & Research/OYSUP Project/oysup_self.csv")
```

First, restructuring data:

```
purpose_long <- tbl_df(purpose) %>%
  gather(-c(FAMID, SEX2, MEDUC2, MPEDUC2), key = "grade", value = "value") %>%
  separate(grade, into = c("variable", "grade"), sep = "_", convert = T) %>%
  spread(variable, value)

purpose_long_elem <- purpose_long %>%
  filter(grade != 21)
```

1. Run linear models on all of your subjects (a basic regression). What is the average intercept, the average slope?

```
model1 <- lm(cbmom ~ grade, data = purpose_long)
summary(model1)

##
## Call:
## lm(formula = cbmom ~ grade, data = purpose_long)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6754 -1.3599 -0.4388  0.6401  5.4035
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.28108    0.07657  16.731  < 2e-16 ***
## grade        0.07886    0.01987   3.969  7.4e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.38 on 3032 degrees of freedom
## (3410 observations deleted due to missingness)
## Multiple R-squared:  0.005167, Adjusted R-squared:  0.004839
## F-statistic: 15.75 on 1 and 3032 DF, p-value: 7.4e-05
```

Average intercept = 1.28

Average slope = .08

## 2. Now run a mlm/lmer model with only a random intercept.

```
model2 <- lmer(cbmom ~ 1 + (1 | FAMID), data = purpose_long)
summary(model2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: cbmom ~ 1 + (1 | FAMID)
## Data: purpose_long
##
## REML criterion at convergence: 10280.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.4057 -0.6586 -0.0506  0.5544  4.0774
##
## Random effects:
## Groups   Name      Variance Std.Dev.
## FAMID    (Intercept) 0.6328   0.7955
## Residual                1.2939   1.1375
## Number of obs: 3034, groups: FAMID, 1063
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  1.60288    0.03302   48.54
```

## 3. What is the ICC?

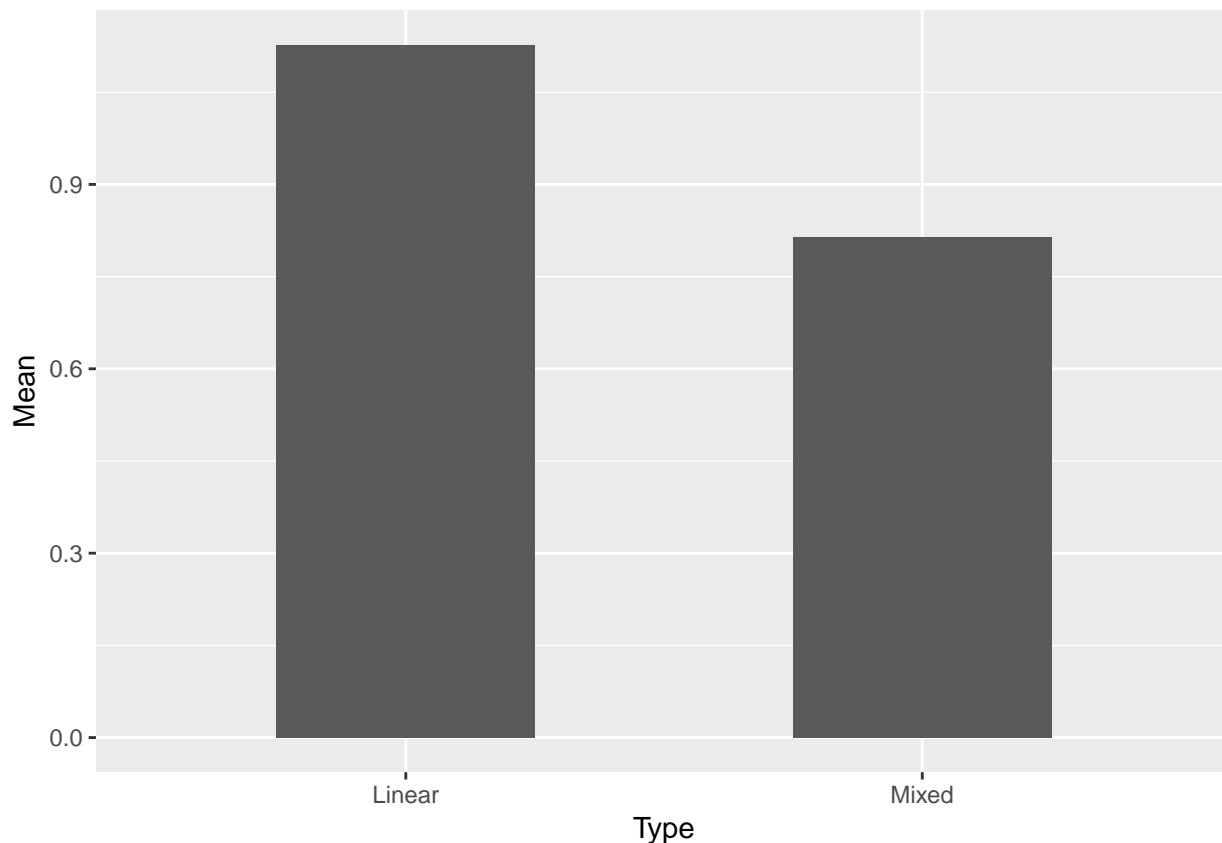
ICC = % between- vs. within-person variance

variance by ID / variance by ID + residual variance

$.64 / (.64 + 1.30) = .33$

What does residual variance look like compared to linear model? Create a graph to show this effect.

```
model1.aug <- augment(model1)
model2.aug <- augment(model2)
mod1_resid <- abs(model1.aug$.resid)
mod2_resid <- abs(model2.aug$.resid)
mean_mod1 <- mean(mod1_resid)
mean_mod2 <- mean(mod2_resid)
resid_df <- data.frame("Type" = c("Linear", "Mixed"),
                      "Mean" = c(mean_mod1, mean_mod2))
resid_plot <- ggplot(resid_df, aes(x = Type, y = Mean)) +
  geom_col(width = .5)
resid_plot
```



Absolute value of the average residual variance is lower in the mixed model, since we're accounting for individual-level, random effects.

**3. Introduce a fixed slope term. What is the difference in terms of the fixed effects estimates between this estimate and the previous? Of the residual standard error? Create a graph to show both fixed effects estimates and the CIs around them.**

```
model3 <- lmer(cbmom ~ grade + (1 | FAMID), data = purpose_long)
summary(model3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: cbmom ~ grade + (1 | FAMID)
## Data: purpose_long
##
## REML criterion at convergence: 10269.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.4634 -0.6873 -0.1045  0.5563  4.0622
##
## Random effects:
## Groups   Name      Variance Std.Dev.
## FAMID    (Intercept) 0.6296   0.7935
## Residual                    1.2867   1.1343
```

```
## Number of obs: 3034, groups:  FAMID, 1063
##
## Fixed effects:
##           Estimate Std. Error t value
## (Intercept)  1.32026    0.07482  17.647
## grade        0.07467    0.01775   4.207
##
## Correlation of Fixed Effects:
##      (Intr)
## grade -0.898
```

Fixed effect estimate for intercept decreased from 1.60 to 1.32.

#### 4. Run an additional model with a random slope. How does this change compare to the previous model? Should you keep the random slope or not?

```
model4 <- lmer(cbmom ~ grade + (1 + grade | FAMID), data = purpose_long)
summary(model4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: cbmom ~ grade + (1 + grade | FAMID)
## Data: purpose_long
##
## REML criterion at convergence: 10243.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.7435 -0.6404 -0.0741  0.5636  4.0278
##
## Random effects:
## Groups   Name                Variance Std.Dev. Corr
## FAMID    (Intercept)  0.65308  0.8081
##          grade        0.04358  0.2087  -0.48
## Residual                1.19431  1.0928
## Number of obs: 3034, groups:  FAMID, 1063
##
## Fixed effects:
##           Estimate Std. Error t value
## (Intercept)  1.30324    0.07316  17.813
## grade        0.07852    0.01844   4.257
##
## Correlation of Fixed Effects:
##      (Intr)
## grade -0.894
```

#### 5. Interpret the correlation between the slope and the intercept.

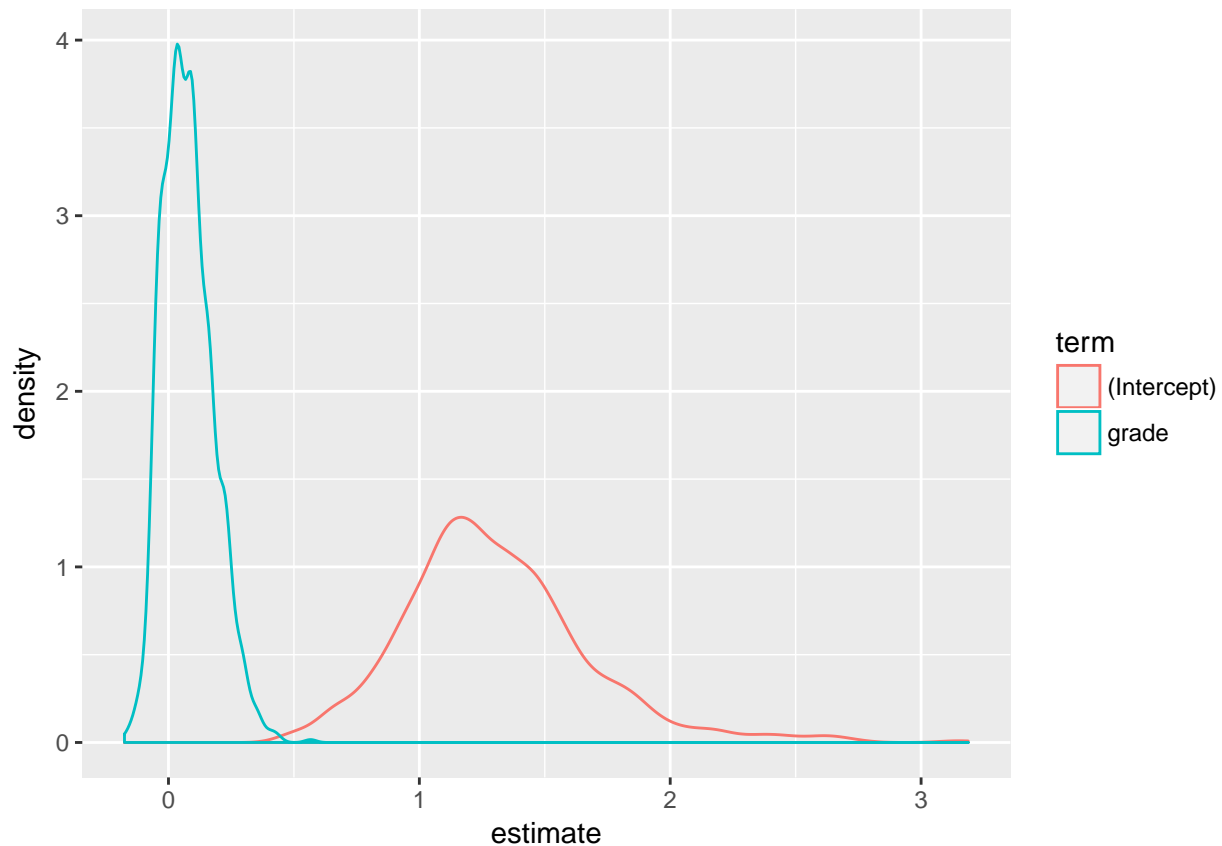
Correlation between slope and intercept = -.89.

6. Create a density plot of the random effects from your final model.

```
random_params <- tidy(model4, effect = "ran_modes")

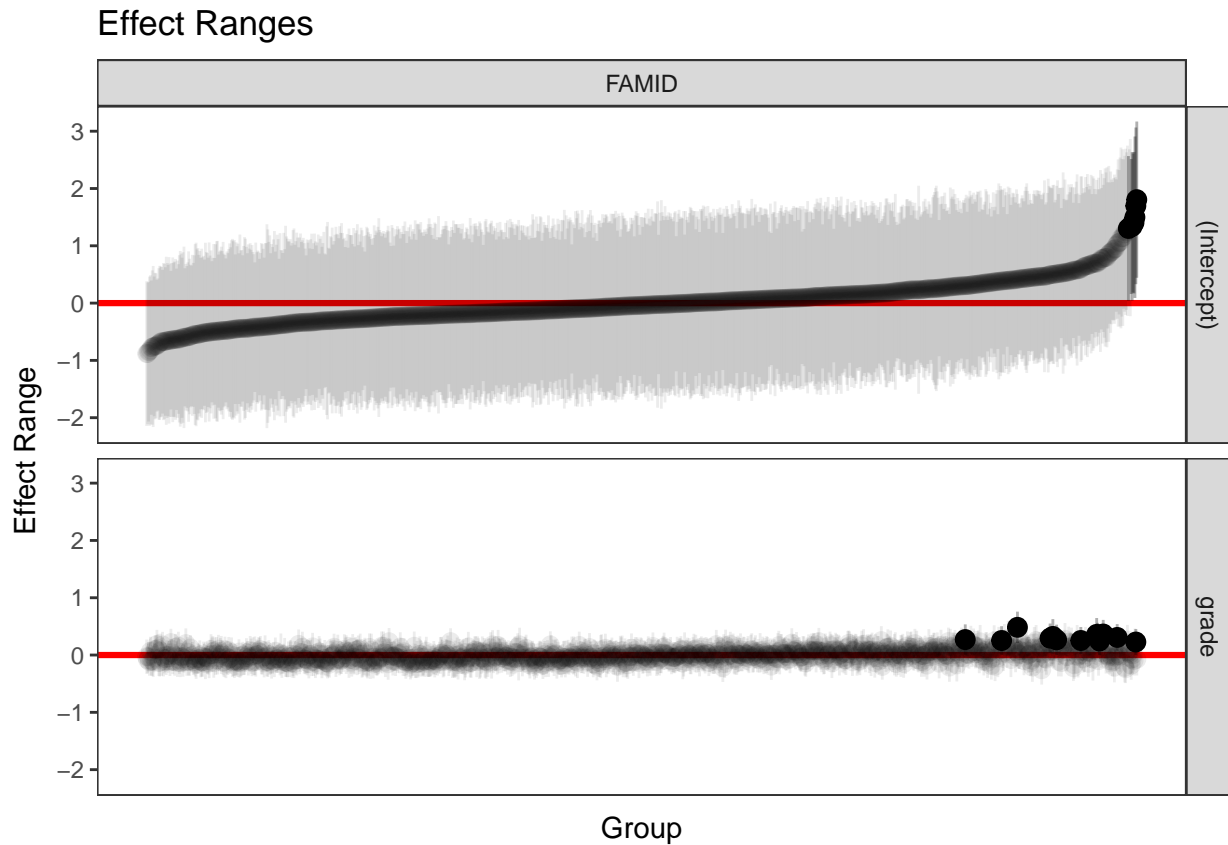
raneff_plot <- ggplot(random_params, aes(x = estimate, color = term)) +
  geom_density()

raneff_plot
```



7. Create a caterpillar plot of the random effects. Is there any person that seems odd in terms of a large standard errors around intercept and slope estimates?

```
re.sim <- RESim(model4)
plotRESim(re.sim)
```



8. Create a plot of the trajectory, along with a spaghetti plot of each person's individual slope. Set the alpha level (transparency) on the individual slopes to make them easier to see.

```
predict <- predictInterval(merMod = model4, newdata = purpose_long_elem,
                           level = 0.9, n.sims = 100,
                           stat = "median", include.resid.var = TRUE)
growth_df <- cbind(purpose_long_elem, predict$fit)

growth_plot <- ggplot(purpose_long_elem, aes(x = grade, y = predict$fit)) +
  geom_line(aes(group = FAMID), alpha = .2) +
  stat_smooth(method = lm) +
  theme_bw()+
  xlab("Grade")+
  ylab("Child-reported conflict with mother")
growth_plot
```

