Homework 3

Leah Schultz

9/28/2017

Chapter 4: Conditional Predictors

```
library(dplyr)
library(lme4)
library(ggplot2)
library(merTools)
library(sjPlot)
library(broom)
library(tidyr)
oysup <- read.csv("~/1-descriptives-and-graphs-leahschultz/oysup_teacher_self.csv")
purpose <- read.csv("~/Dropbox/Lab & Research/OYSUP Project/oysup_self.csv")
oysup <- oysup %>%
  dplyr::select(FAMID, neuro_7s:neuro_10s)
dems <- purpose %>%
  dplyr::select(SEX2, MPEDUC2)
oysup <- cbind(oysup, dems)
```

First, restructuring data:

```
oysup_long <- tbl_df(oysup) %>%
  gather(c(neuro_7s:neuro_10s), key = "grade", value = "value") %>%
  separate(grade, into = c("variable", "grade"), sep = "_", convert = T) %>%
  separate(grade, into = c("grade", "delete"), sep = "s") %>%
  mutate(grade = as.numeric(grade)) %>%
  dplyr::select(-delete) %>%
  spread(variable, value)
oysup_long
```

```
## # A tibble: 4,296 x 5
##    FAMID  SEX2 MPEDUC2 grade neuro
##  * <int> <int>   <int> <dbl> <dbl>
##  1  1001     2       3     7    NA
##  2  1001     2       3     8    NA
##  3  1001     2       3     9   3.5
##  4  1001     2       3    10   5.0
##  5  1002     2       3     7   3.5
##  6  1002     2       3     8   3.5
##  7  1002     2       3     9   2.0
##  8  1002     2       3    10   2.5
##  9  1003     1       3     7    NA
## 10  1003     1       3     8   4.0
## # ... with 4,286 more rows
```

1. Run a series of models using a time-invariant nominal covariate, a) where the covariate only predicts the intercept b) predicts both intercept and slope c) is rescaled eg centering. For all models, how does your model change from model to model? What is your final model?

```
## Getting a subset so that the models are on equivalent datasets:
oysup_long2 <- subset(oysup_long, subset = !is.na(SEX2))

model_null <- lmer(neuro ~ grade + (1 | FAMID), data = oysup_long2)

model1 <- lmer(neuro ~ grade + SEX2 + (1 | FAMID), data = oysup_long2)
summary(model1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: neuro ~ grade + SEX2 + (1 | FAMID)
##    Data: oysup_long2
##
## REML criterion at convergence: 8270.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5935 -0.5650  0.0324  0.6012  2.8594
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  FAMID    (Intercept) 0.4655   0.6823
##  Residual             0.5996   0.7744
## Number of obs: 3051, groups:  FAMID, 934
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  2.91460    0.14631  19.921
## grade       -0.04320    0.01363  -3.168
## SEX2         0.42993    0.05325   8.074
##
## Correlation of Fixed Effects:
##       (Intr) grade
## grade -0.816
## SEX2  -0.552  0.004
```

```
anova(model_null, model1)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: oysup_long2
## Models:
## model_null: neuro ~ grade + (1 | FAMID)
## model1: neuro ~ grade + SEX2 + (1 | FAMID)
##            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_null  4 8325.2 8349.3 -4158.6   8317.2
## model1      5 8264.2 8294.3 -4127.1   8254.2 63.006      1  2.061e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When the covariate predicts the intercept, we can see that gender accounts for some of the differences in initial neuroticism (which is verified by the improved fit in the model when gender is included). Specifically, female students rated themselves as more neurotic on average than male students (by .43 points on a scale from 1 to 5), across grades.

```
model2 <- lmer(neuro ~ grade*SEX2 + (1 | FAMID), data = oysup_long2)
summary(model2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: neuro ~ grade * SEX2 + (1 | FAMID)
##    Data: oysup_long2
##
## REML criterion at convergence: 8272.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5808 -0.5687  0.0409  0.5984  2.9067
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  FAMID    (Intercept) 0.4667   0.6832
##  Residual             0.5988   0.7738
## Number of obs: 3051, groups:  FAMID, 934
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  3.52661    0.39056   9.030
## grade       -0.11327    0.04365  -2.595
## SEX2         0.02762    0.24399   0.113
## grade:SEX2   0.04608    0.02727   1.690
##
## Correlation of Fixed Effects:
##            (Intr) grade  SEX2
## grade      -0.976
## SEX2       -0.950  0.927
## grade:SEX2  0.927 -0.950 -0.976
```

```
anova(model1, model2)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: oysup_long2
## Models:
## model1: neuro ~ grade + SEX2 + (1 | FAMID)
## model2: neuro ~ grade * SEX2 + (1 | FAMID)
##        Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## model1  5 8264.2 8294.3 -4127.1   8254.2
## model2  6 8263.4 8299.5 -4125.7   8251.4 2.851      1    0.09132 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When the covariate predicts the slope as well as the intercept and we get an interaction between time and gender, we can see that gender no longer accounts for differences in initial neuroticism. In addition, the interaction doesn't seem to add anything to the model. It doesn't seem that gender affects the degree to which adolescents change in neuroticism, even if it affects their initial levels.

```
SEX2_Z <- scale(oysup_long2$SEX2, center = T)
```

```r
model_null <- lmer(neuro ~ grade + (1 | FAMID), data = oysup_long2)

model3 <- lmer(neuro ~ grade + SEX2_Z + (1 | FAMID), data = oysup_long2)
summary(model3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: neuro ~ grade + SEX2_Z + (1 | FAMID)
##    Data: oysup_long2
##
## REML criterion at convergence: 8271.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5935 -0.5650  0.0324  0.6012  2.8594
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  FAMID    (Intercept) 0.4655   0.6823
##  Residual             0.5996   0.7744
## Number of obs: 3051, groups:  FAMID, 934
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  3.55929    0.12197  29.183
## grade       -0.04320    0.01363  -3.168
## SEX2_Z       0.21499    0.02663   8.074
##
## Correlation of Fixed Effects:
##        (Intr) grade
## grade  -0.976
## SEX2_Z -0.008  0.004
```

```r
anova(model_null, model3)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: oysup_long2
## Models:
## model_null: neuro ~ grade + (1 | FAMID)
## model3: neuro ~ grade + SEX2_Z + (1 | FAMID)
##            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_null  4 8325.2 8349.3 -4158.6   8317.2
## model3      5 8264.2 8294.3 -4127.1   8254.2 63.006      1  2.061e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
model4 <- lmer(neuro ~ grade*SEX2_Z + (1 | FAMID), data = oysup_long2)
summary(model4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: neuro ~ grade * SEX2_Z + (1 | FAMID)
##    Data: oysup_long2
##
## REML criterion at convergence: 8275.7
##
## Scaled residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -3.5808 -0.5687  0.0409  0.5984  2.9067
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  FAMID    (Intercept) 0.4667   0.6832
##  Residual             0.5988   0.7738
## Number of obs: 3051, groups:  FAMID, 934
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)   3.56802    0.12200  29.246
## grade        -0.04417    0.01364  -3.239
## SEX2_Z        0.01381    0.12201   0.113
## grade:SEX2_Z  0.02304    0.01364   1.690
##
## Correlation of Fixed Effects:
##             (Intr) grade  SEX2_Z
## grade       -0.976
## SEX2_Z      -0.043  0.042
## grad:SEX2_Z  0.042 -0.042 -0.976
```

```
anova(model3, model4)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: oysup_long2
## Models:
## model3: neuro ~ grade + SEX2_Z + (1 | FAMID)
## model4: neuro ~ grade * SEX2_Z + (1 | FAMID)
##        Df    AIC    BIC  logLik deviance Chisq Chi Df Pr(>Chisq)
## model3  5 8264.2 8294.3 -4127.1   8254.2
## model4  6 8263.4 8299.5 -4125.7   8251.4 2.851      1    0.09132 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When I standardized the sex variable, the estimate was cut in half, to a difference of .21 units in neuroticism between men and women. Previously, the data was coded as 1 and 2 instead of as 0 and 1.

My final model is model C, the model predicting the intercept but not the slope, using the standardized sex variable.

2. Introduce a time-invariant continuous covariate and run models a-c from #1.

```
oysup_long3 <- subset(oysup_long, subset = !is.na(MPEDUC2))

model_null <- lmer(neuro ~ grade + (1 | FAMID), data = oysup_long3)

#A)
model1 <- lmer(neuro ~ grade + MPEDUC2 + (1 | FAMID), data = oysup_long3)
summary(model1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: neuro ~ grade + MPEDUC2 + (1 | FAMID)
##    Data: oysup_long3
```

```
##
## REML criterion at convergence: 5973.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5672 -0.5653  0.0419  0.5946  2.8844
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  FAMID    (Intercept) 0.4807   0.6933
##  Residual             0.6031   0.7766
## Number of obs: 2194, groups:  FAMID, 663
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  3.85458    0.15655  24.621
## grade       -0.05398    0.01605  -3.364
## MPEDUC2     -0.07288    0.01853  -3.933
##
## Correlation of Fixed Effects:
##         (Intr) grade
## grade   -0.890
## MPEDUC2 -0.401 -0.007
```

```r
anova(model_null, model1)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: oysup_long3
## Models:
## model_null: neuro ~ grade + (1 | FAMID)
## model1: neuro ~ grade + MPEDUC2 + (1 | FAMID)
##            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_null  4 5978.9 6001.7 -2985.4   5970.9
## model1      5 5965.5 5994.0 -2977.8   5955.5 15.342      1  8.971e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A) As students' fathers' education level increases by 1 unit, the students' self-reported neuroticism decreases by .07 units.

```r
#B)
model2 <- lmer(neuro ~ grade*MPEDUC2 + (1 | FAMID), data = oysup_long3)
summary(model2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: neuro ~ grade * MPEDUC2 + (1 | FAMID)
##    Data: oysup_long3
##
## REML criterion at convergence: 5979.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5496 -0.5613  0.0386  0.5935  2.9132
##
## Random effects:
```

```
##  Groups   Name        Variance Std.Dev.
##  FAMID    (Intercept) 0.4803   0.6931
##  Residual             0.6031   0.7766
## Number of obs: 2194, groups:  FAMID, 663
##
## Fixed effects:
##                Estimate Std. Error t value
## (Intercept)    3.552666   0.317931  11.174
## grade         -0.019237   0.035657  -0.540
## MPEDUC2        0.014653   0.082337   0.178
## grade:MPEDUC2 -0.010063   0.009224  -1.091
##
## Correlation of Fixed Effects:
##             (Intr) grade  MPEDUC
## grade       -0.975
## MPEDUC2     -0.892  0.869
## grd:MPEDUC2  0.870 -0.893 -0.974
```

```r
anova(model1, model2)
```

```
## refitting model(s) with ML (instead of REML)

## Data: oysup_long3
## Models:
## model1: neuro ~ grade + MPEDUC2 + (1 | FAMID)
## model2: neuro ~ grade * MPEDUC2 + (1 | FAMID)
##        Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model1  5 5965.5 5994.0 -2977.8   5955.5
## model2  6 5966.3 6000.5 -2977.2   5954.3 1.1922      1     0.2749
```

B) Fathers' level of education does not affect the degree to which students change in self-reported neuroticism over their adolescence.

```r
#C)
MPEDUC2_Z <- scale(oysup_long3$MPEDUC2, center = T, scale = F)

model3 <- lmer(neuro ~ grade + MPEDUC2_Z + (1 | FAMID), data = oysup_long3)
summary(model3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: neuro ~ grade + MPEDUC2_Z + (1 | FAMID)
##    Data: oysup_long3
##
## REML criterion at convergence: 5973.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5672 -0.5653  0.0419  0.5946  2.8844
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  FAMID    (Intercept) 0.4807   0.6933
##  Residual             0.6031   0.7766
## Number of obs: 2194, groups:  FAMID, 663
##
## Fixed effects:
```

```
##              Estimate Std. Error t value
## (Intercept)  3.60485    0.14342  25.135
## grade       -0.05398    0.01605  -3.364
## MPEDUC2_Z   -0.07288    0.01853  -3.933
##
## Correlation of Fixed Effects:
##            (Intr) grade
## grade     -0.975
## MPEDUC2_Z  0.005 -0.007
```

```
anova(model_null, model3)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: oysup_long3
## Models:
## model_null: neuro ~ grade + (1 | FAMID)
## model3: neuro ~ grade + MPEDUC2_Z + (1 | FAMID)
##            Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model_null  4 5978.9 6001.7 -2985.4   5970.9
## model3      5 5965.5 5994.0 -2977.8   5955.5 15.342      1  8.971e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
model4 <- lmer(neuro ~ grade*MPEDUC2_Z + (1 | FAMID), data = oysup_long3)
summary(model4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: neuro ~ grade * MPEDUC2_Z + (1 | FAMID)
##    Data: oysup_long3
##
## REML criterion at convergence: 5979.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5496 -0.5613  0.0386  0.5935  2.9132
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  FAMID    (Intercept) 0.4803   0.6931
##  Residual             0.6031   0.7766
## Number of obs: 2194, groups:  FAMID, 663
##
## Fixed effects:
##                   Estimate Std. Error t value
## (Intercept)       3.602878   0.143433  25.119
## grade            -0.053722   0.016050  -3.347
## MPEDUC2_Z         0.014653   0.082337   0.178
## grade:MPEDUC2_Z -0.010063   0.009224  -1.091
##
## Correlation of Fixed Effects:
##              (Intr) grade  MPEDUC
## grade        -0.975
## MPEDUC2_Z    -0.011  0.013
## g:MPEDUC2_Z  0.013 -0.015 -0.974
```

```r
anova(model3, model4)
```

```
## refitting model(s) with ML (instead of REML)

## Data: oysup_long3
## Models:
## model3: neuro ~ grade + MPEDUC2_Z + (1 | FAMID)
## model4: neuro ~ grade * MPEDUC2_Z + (1 | FAMID)
##        Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## model3  5 5965.5 5994.0 -2977.8   5955.5
## model4  6 5966.3 6000.5 -2977.2   5954.3 1.1922      1     0.2749
```

C) Though the estimate for the grade variable doesn't change, the estimate for a one unit increase in father's education becomes a decrease of .13 points on neuroticism, compared to .07. The interaction term still does not contribute anything meaningful to the model.

3. Graph both of your final models for the continuous and nominal models above.

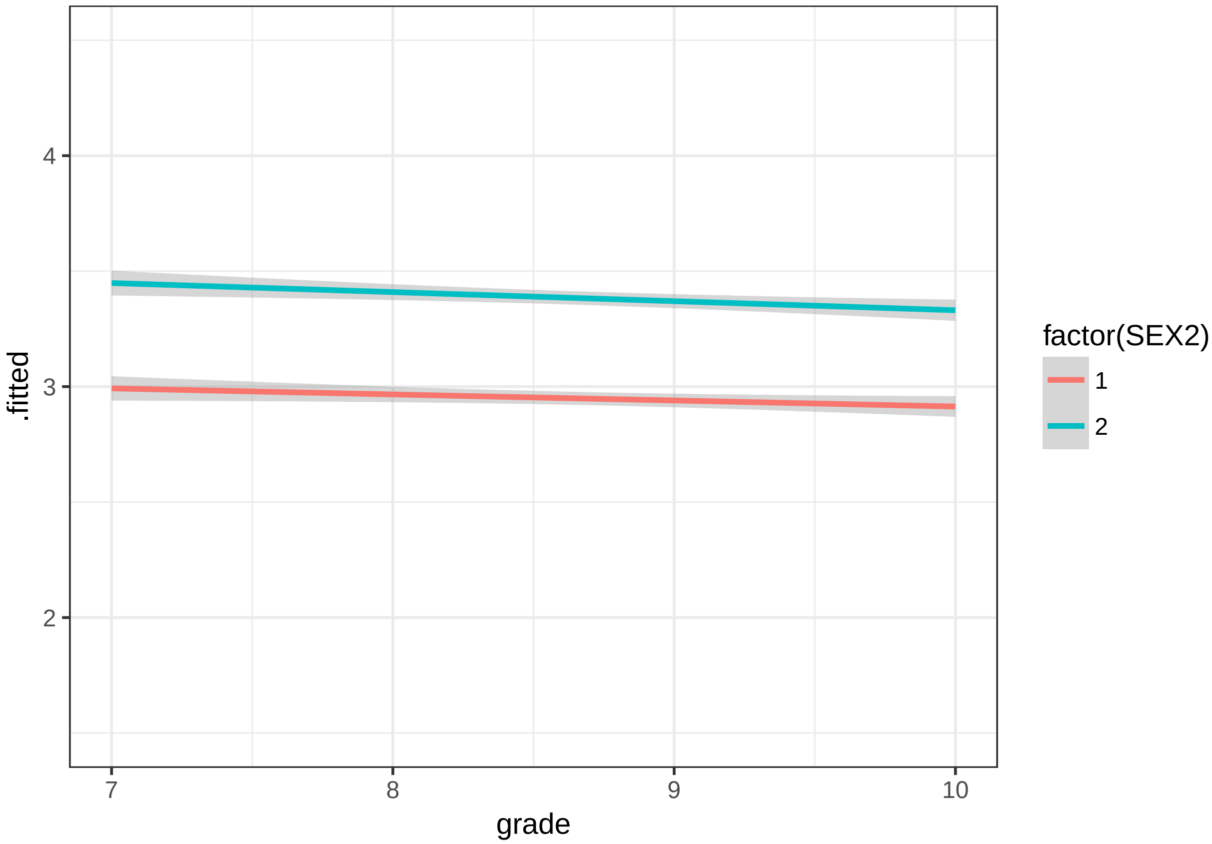
```r
oysup_long2$SEX2 <- as.factor(oysup_long2$SEX2)

model3 <- lmer(neuro ~ grade + SEX2 + (1 | FAMID), data = oysup_long2)
summary(model3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: neuro ~ grade + SEX2 + (1 | FAMID)
##    Data: oysup_long2
##
## REML criterion at convergence: 8270.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5935 -0.5650  0.0324  0.6012  2.8594
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  FAMID    (Intercept) 0.4655   0.6823
##  Residual             0.5996   0.7744
## Number of obs: 3051, groups:  FAMID, 934
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  3.34454    0.12504  26.748
## grade       -0.04320    0.01363  -3.168
## SEX22        0.42993    0.05325   8.074
##
## Correlation of Fixed Effects:
##       (Intr) grade
## grade -0.953
## SEX22 -0.220  0.004
```

```r
library(broom)
model3_aug <- augment(model3)
nominal_plot <- ggplot(model3_aug, aes(x = grade, y = .fitted, color=factor(SEX2))) +
  stat_smooth(aes(group = SEX2), method="lm") +
```

```
  ylim(1.5,4.5) +
  theme_bw()
nominal_plot
```



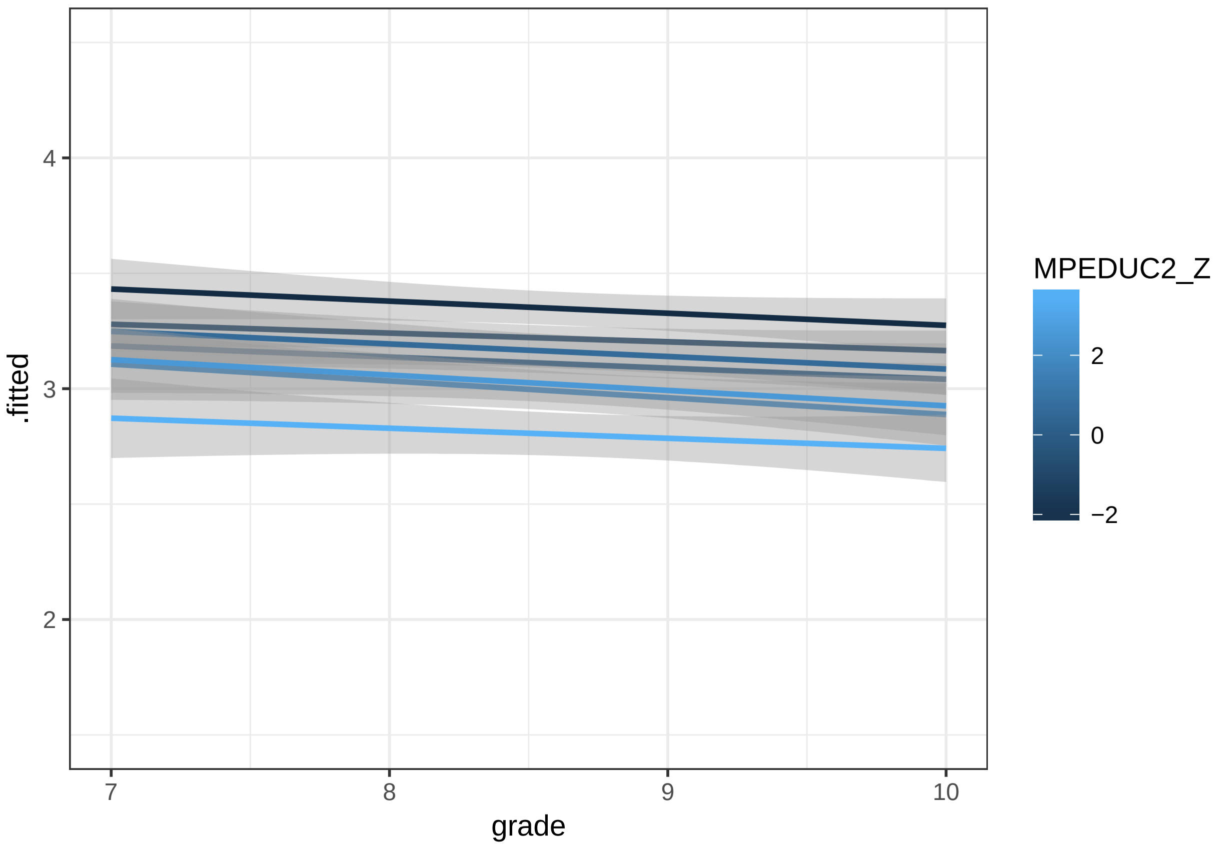

```
model4 <- lmer(neuro ~ grade + MPEDUC2_Z + (1 | FAMID), data = oysup_long3)
summary(model4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: neuro ~ grade + MPEDUC2_Z + (1 | FAMID)
##    Data: oysup_long3
##
## REML criterion at convergence: 5973.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5672 -0.5653  0.0419  0.5946  2.8844
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  FAMID    (Intercept) 0.4807   0.6933
##  Residual             0.6031   0.7766
## Number of obs: 2194, groups:  FAMID, 663
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  3.60485    0.14342  25.135
## grade       -0.05398    0.01605  -3.364
```

```
## MPEDUC2_Z   -0.07288    0.01853  -3.933
##
## Correlation of Fixed Effects:
##           (Intr) grade
## grade     -0.975
## MPEDUC2_Z  0.005 -0.007
```

```
model4_aug <- augment(model4)

continuous_plot <- ggplot(model4_aug, aes(x = grade, y = .fitted)) +
  stat_smooth(aes(color = MPEDUC2_Z, group = MPEDUC2_Z),method="lm")+
  ylim(1.5,4.5)+
  theme_bw()
continuous_plot
```



4. Calculate confidence intervals around your estimates for your final models.

```
confint(model3, parm = c("grade", "SEX22"))
```

```
## Computing profile confidence intervals ...
```

```
##             2.5 %      97.5 %
## grade -0.06991332 -0.01645481
## SEX22  0.32555477  0.53428102
```

```
confint(model4, parm = c("grade", "MPEDUC2_Z"))
```

```
## Computing profile confidence intervals ...

##                  2.5 %      97.5 %
## grade     -0.08543471 -0.02251542
## MPEDUC2_Z -0.10919247 -0.03656621
```

5. Include both types of covariates in a single model. How does your interpretation of parameters change?

```
oysup_long5 <- subset(oysup_long, subset = !is.na(MPEDUC2) & !is.na(SEX2))
SEX2_Z <- scale(oysup_long5$SEX2, center = T, scale = F)
MPEDUC2_Z <- scale(oysup_long5$MPEDUC2, center = T, scale = F)
oysup_long5 <- cbind(oysup_long5, SEX2_Z, MPEDUC2_Z)
model5 <- lmer(neuro ~ grade + MPEDUC2_Z + SEX2_Z + (1 | FAMID), data = oysup_long5)
summary(model5)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: neuro ~ grade + MPEDUC2_Z + SEX2_Z + (1 | FAMID)
##    Data: oysup_long5
##
## REML criterion at convergence: 5846.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6394 -0.5655  0.0295  0.5850  2.8318
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  FAMID    (Intercept) 0.4500   0.6708
##  Residual             0.5957   0.7718
## Number of obs: 2165, groups:  FAMID, 655
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  3.61177    0.14347  25.174
## grade       -0.05602    0.01606  -3.487
## MPEDUC2_Z   -0.07073    0.01824  -3.878
## SEX2_Z       0.37617    0.06268   6.002
##
## Correlation of Fixed Effects:
##           (Intr) grade  MPEDUC
## grade     -0.976
## MPEDUC2_Z  0.005 -0.007
## SEX2_Z    -0.006  0.000  0.032
```

```
model5_aug <- augment(model5)
```

The effects of these variables don't seem to overlap very much, as their effects are preserved relative to the models in which they are each alone. However, now we can say that, holding father's education level constant, female students rate themselves higher on neuroticism than males do. And of course, sex of the student remaining constant, students who come from families where the father is more highly educated tend to rate themselves as lower on neuroticism.

6. If you have one available, introduce a time-varying covariate.