Week 5 Homework

Molly Farry-Thorn

October 3, 2017

Contents

1. Run a series of models using a time-invariant nominal covariate. For all models, how does your model change from model to model. What is your final model?

a) where the covariate only predicts the intercept

```
# time invariant covariate that predicts the intecept but not slope
children$Child.Gender <- relevel(children$Child.Gender, ref = "Male")

mod1a <- lmer(Utterances.with.Letters ~ Time.c + Child.Gender + (Time.c|Subject), data = children)
summary(mod1a)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Utterances.with.Letters ~ Time.c + Child.Gender + (Time.c | Subject)
##    Data: children
##
## REML criterion at convergence: 5394
##
```

```
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6682 -0.2603 -0.1622 -0.0625  9.9127
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept)  28.43    5.332
##           Time.c       14.52    3.810   1.00
##  Residual             207.76   14.414
## Number of obs: 652, groups:  Subject, 55
##
## Fixed effects:
##                    Estimate Std. Error t value
## (Intercept)          6.0681     1.1584   5.238
## Time.c               2.6346     0.7096   3.713
## Child.GenderFemale  -0.6013     1.5066  -0.399
##
## Correlation of Fixed Effects:
##             (Intr) Time.c
## Time.c       0.448
## Chld.GndrFm -0.614  0.001
```

Fixed Effects

Intercept: 6.07; mean number of utterances with letters for male children at the mean age

Time: 2.63; increase in number of utterances with letters every year

Child.Gender: -0.60; difference between males and females at mean age, females start lower

b) predicts both intercept and slope

```r
# time invariant predictor for the intercept AND slopes
mod1b <- lmer(Utterances.with.Letters ~ Time.c + Child.Gender + Time.c*Child.Gender + (Time.c|Subject),
summary(mod1b)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Utterances.with.Letters ~ Time.c + Child.Gender + Time.c * Child.Gender +
##     (Time.c | Subject)
##    Data: children
##
## REML criterion at convergence: 5389.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7199 -0.2644 -0.1659 -0.0529  9.8806
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept)  28.24    5.314
##           Time.c       14.22    3.771   1.00
##  Residual             207.70   14.412
## Number of obs: 652, groups:  Subject, 55
```

```
##
## Fixed effects:
##                            Estimate Std. Error t value
## (Intercept)                  6.7060     1.2562   5.338
## Time.c                       3.5066     0.9733   3.603
## Child.GenderFemale          -1.9484     1.8273  -1.066
## Time.c:Child.GenderFemale   -1.8385     1.4132  -1.301
##
## Correlation of Fixed Effects:
##             (Intr) Time.c Chl.GF
## Time.c       0.565
## Chld.GndrFm -0.687 -0.389
## Tm.c:Chl.GF -0.389 -0.689  0.565
```

Fixed Effects

Intercept: 6.71; mean number of utterances with letters for male children at mean age

Time: 3.51; increase in number of utterances with letters every year, for males

Child.Gender: -1.95; difference between males and females at mean age, females are lower

Time:Child.Gender: -1.84; difference in the slopes between males and females, the effect of age is smaller in females

c) is rescaled (e.g. centering).

```
# changing dummy coding such that reference group becomes females
children$Child.Gender <- relevel(children$Child.Gender, ref = "Female")
mod1c <- lmer(Utterances.with.Letters ~ Time.c + Child.Gender + Time.c*Child.Gender + (Time.c|Subject),
summary(mod1c)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Utterances.with.Letters ~ Time.c + Child.Gender + Time.c * Child.Gender +
##     (Time.c | Subject)
##    Data: children
##
## REML criterion at convergence: 5389.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7199 -0.2644 -0.1659 -0.0529  9.8806
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept)  28.24    5.314
##           Time.c       14.22    3.771   1.00
##  Residual             207.70   14.412
## Number of obs: 652, groups:  Subject, 55
##
## Fixed effects:
##                          Estimate Std. Error t value
## (Intercept)                 4.758      1.327   3.585
## Time.c                      1.668      1.025   1.628
```

```
## Child.GenderMale           1.948      1.827   1.066
## Time.c:Child.GenderMale    1.839      1.413   1.301
##
## Correlation of Fixed Effects:
##             (Intr) Time.c Chl.GM
## Time.c       0.565
## Chld.GndrMl -0.726 -0.410
## Tm.c:Chl.GM -0.409 -0.725  0.565
```

Fixed Effects

Intercept: 4.76; mean number of utterances with letters for female children at mean age

Time: 1.67; increase in number of utterances with letters every year, for females

Child.Gender: 1.95; difference between males and females at mean age, males start higher. This is the same difference we found in mod1b, just in the opposite direction.

Time:Child.Gender: 1.84; difference in the slopes between males and females, the effect of age is larger in males. This is the same difference we found in mod1b, just in the opposite direction.

```
anova(mod1a,mod1b)
```

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

```
## Data: children
## Models:
## mod1a: Utterances.with.Letters ~ Time.c + Child.Gender + (Time.c | Subject)
## mod1b: Utterances.with.Letters ~ Time.c + Child.Gender + Time.c * Child.Gender +
## mod1b:     (Time.c | Subject)
##       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod1a  7 5413.0 5444.4 -2699.5   5399.0
## mod1b  8 5413.3 5449.2 -2698.7   5397.3 1.7164      1     0.1902
```

The likelihood ratio test suggests that simpler model, where covariate only predicts intercept, is preferred.

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

```
# time-invariant continuous covariate that only predicts the intercept
mod2a <- lmer(Utterances.with.Letters ~ Time.c + SES + (Time.c|Subject), data = children)
summary(mod2a)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Utterances.with.Letters ~ Time.c + SES + (Time.c | Subject)
##    Data: children
##
## REML criterion at convergence: 5394.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6676 -0.2661 -0.1681 -0.0548  9.9259
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept)  27.79    5.271
```

```
##           Time.c       14.53    3.812   1.00
##  Residual             207.67   14.411
## Number of obs: 652, groups:  Subject, 55
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   5.7833     0.9078   6.371
## Time.c        2.6351     0.7097   3.713
## SES           0.7343     0.7556   0.972
##
## Correlation of Fixed Effects:
##        (Intr) Time.c
## Time.c  0.566
## SES    -0.001  0.001
```

Fixed Effects

Intercept: 5.78; mean of number of utterances with letters for children with mean level SES at mean age

Time: 2.64; increase in number of utterances with letters every year, when SES at mean level

SES: 0.73; increase in number of utterances for every 1 unit increase in SES at mean age

```r
# time-invariant continuous covariate predicts the intercept AND slopes
mod2b <- lmer(Utterances.with.Letters ~ Time.c + SES + Time.c*SES + (Time.c|Subject), data = children)
summary(mod2b)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Utterances.with.Letters ~ Time.c + SES + Time.c * SES + (Time.c |
##     Subject)
##    Data: children
##
## REML criterion at convergence: 5392.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6529 -0.2647 -0.1571 -0.0577  9.9348
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept)  27.85    5.277
##           Time.c       14.62    3.824   1.00
##  Residual             207.70   14.412
## Number of obs: 652, groups:  Subject, 55
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   5.7823     0.9084   6.365
## Time.c        2.6331     0.7109   3.704
## SES           1.1902     0.9167   1.298
## Time.c:SES    0.6297     0.7168   0.878
##
## Correlation of Fixed Effects:
##            (Intr) Time.c SES
## Time.c      0.567
## SES        -0.002 -0.001
```

```
## Time.c:SES -0.001 -0.003  0.566
```

Fixed Effects

Intercept: 5.78; mean of number of utterances with letters for children with mean level SES at mean age

Time: 2.63; increase in number of utterances with letters every year, when SES at mean level

SES: 1.19; increase in number of utterances for every 1 unit increase in SES at mean age

Time:SES: 0.63; the change in the relationship in the time slope for every 1 unit increase in SES

My SES variable is already centered.

```
anova(mod2a,mod2b)
```

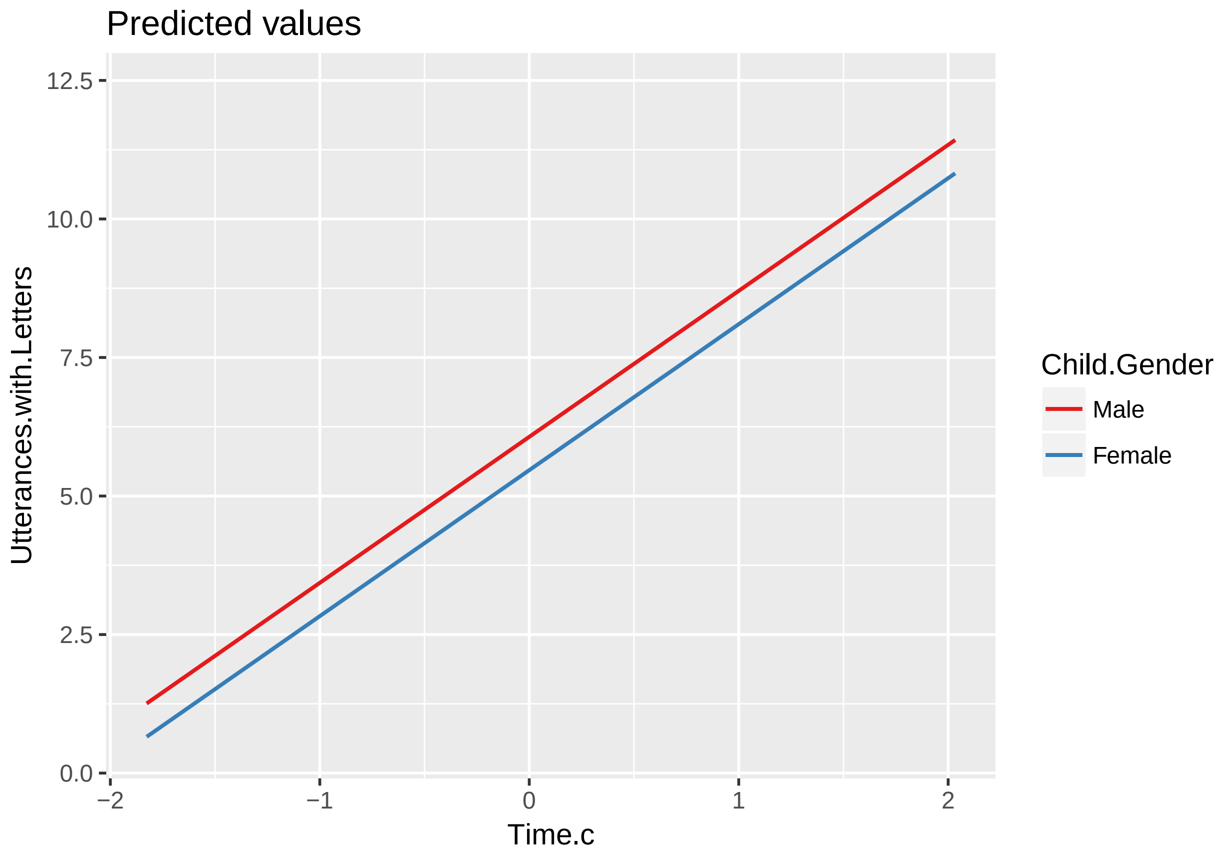
```
## refitting model(s) with ML (instead of REML)
```

```
## Data: children
## Models:
## mod2a: Utterances.with.Letters ~ Time.c + SES + (Time.c | Subject)
## mod2b: Utterances.with.Letters ~ Time.c + SES + Time.c * SES + (Time.c |
## mod2b:     Subject)
##       Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## mod2a  7 5412.3 5443.6 -2699.1   5398.3
## mod2b  8 5413.5 5449.3 -2698.7   5397.5 0.7887      1     0.3745
```

Again, the likelihood ratio test suggests that simpler model, where covariate only predicts intercept, is preferred.

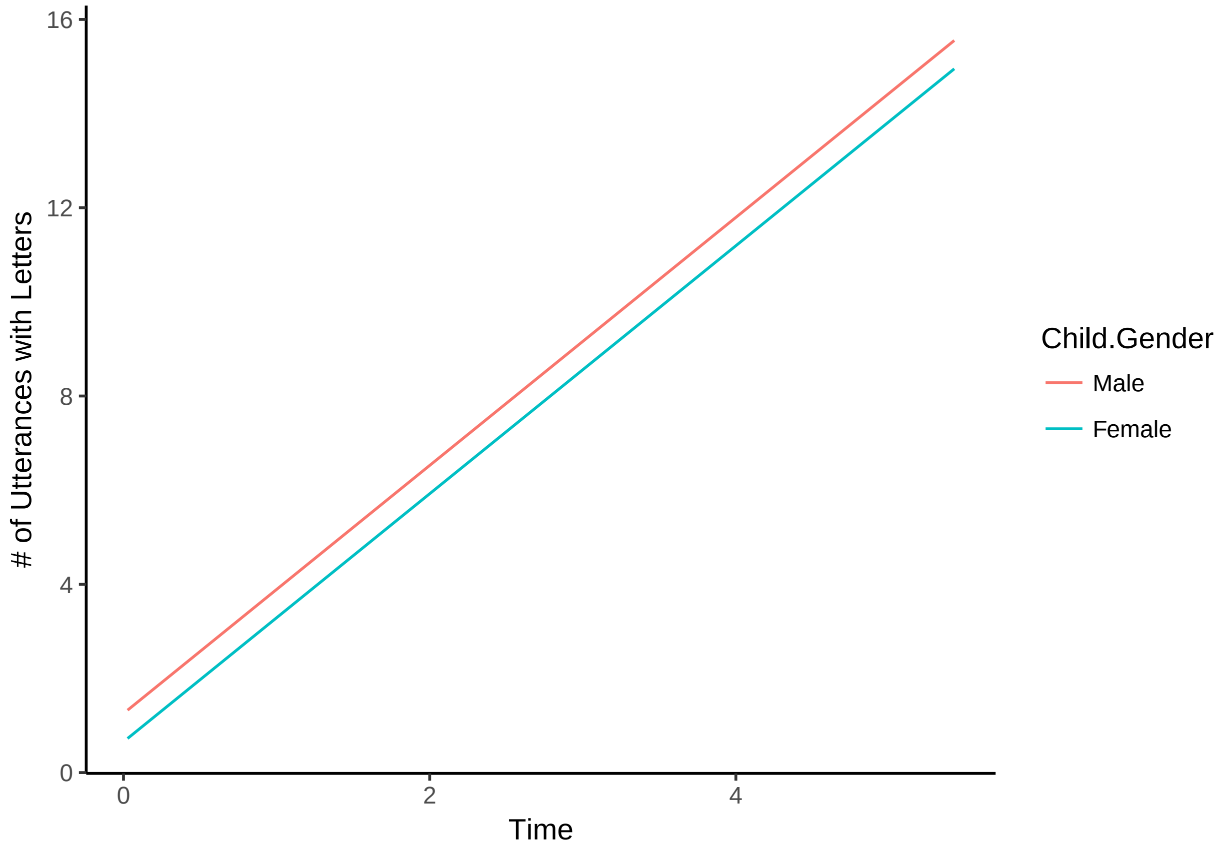
3. Graph both of your final models for the continuous and nominal models above.

```
# Graphing nominal, where gender only predict the intercept
children$Child.Gender <- relevel(children$Child.Gender, ref = "Male")
sjp.lmer(mod1a, type = "pred.fe", var = c("Time.c", "Child.Gender"), facet = FALSE, show.scatter = FALS
```




```
fixed.frame <- data.frame(expand.grid(Time.c = seq(-1.8,4,1.8), Child.Gender = c("Male", "Female"))) %>%
  mutate(pred = predict(mod1a, newdata = ., re.form = NA))
fixed.frame <- fixed.frame %>% mutate(Time.new = Time.c + mean(letters$Time))

nom <- ggplot(aes(x = Time.new, y = pred, color = Child.Gender), data = fixed.frame) +
  geom_line() +
  labs(x = "Time", y = "# of Utterances with Letters") +
  theme_classic()
nom
```

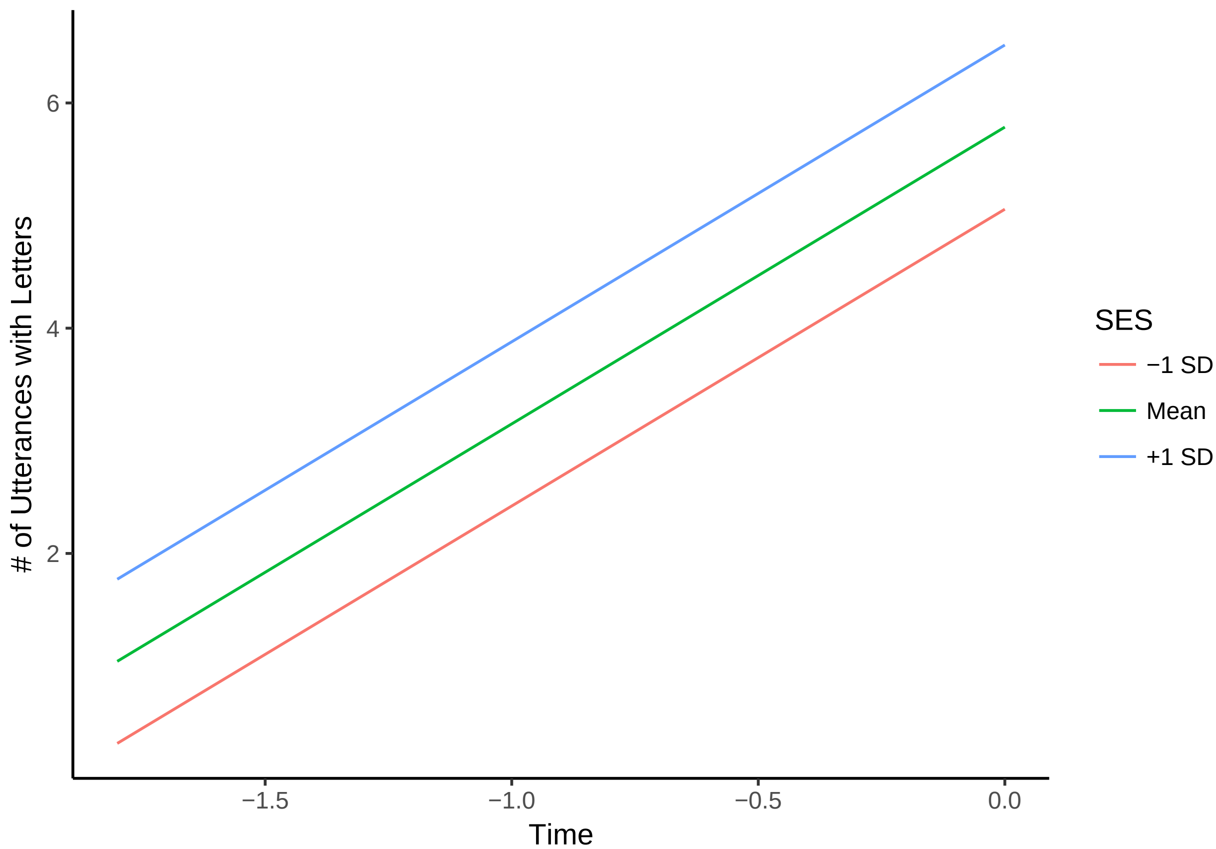



```r
# Graphing continuous, where SES only predicts the intercept
fixed.frame2 <- letters %>%
  summarise(mean = mean(SES, na.rm = T), sd = sd(SES, na.rm = T))
fixed.frame <- fixed.frame %>% mutate(Time.new = Time.c + mean(letters$Time))

fixed.frame2 <- data.frame(
  expand.grid(
    Time.c = seq(-1.8,1,1.8),
    SES = c(fixed.frame2$mean - fixed.frame2$sd,
            fixed.frame2$mean,
            fixed.frame2$mean + fixed.frame2$sd))) %>%
  mutate(pred = predict(mod2a, newdata = ., re.form = NA))

fixed.frame2$SES <- as.factor(fixed.frame2$SES)
levels(fixed.frame2$SES) <- c("-1 SD","Mean", "+1 SD")

con <- ggplot(aes(x = Time.c, y = pred, color = SES), data = fixed.frame2) +
  geom_line() +
  labs(x = "Time", y = "# of Utterances with Letters") +
  theme_classic()
con
```



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

```r
library(broom)
#Confidence intervals around nominal model
mod1a.ci <- confint(mod1a, level = .95, oldNames = F, method = "boot", nsim = 1000)
```

```
## Computing bootstrap confidence intervals ...
```

```r
broom::tidy(mod1a.ci)
```

```
##                        .rownames      X2.5..    X97.5..
## 1          sd_(Intercept)|Subject  3.6937246  7.001716
## 2 cor_Time.c.(Intercept)|Subject  0.7260955  1.000000
## 3               sd_Time.c|Subject  2.5141388  5.097271
## 4                           sigma 13.5874146 15.218372
## 5                     (Intercept)  3.6999322  8.363094
## 6                          Time.c  1.2885404  4.085120
## 7              Child.GenderFemale -3.8293861  2.578967
```

```r
#Confidence intervals around continuous model
mod2a.ci <- confint(mod2a, level = .95, oldNames = F, method = "boot", nsim = 1000)
```

```
## Computing bootstrap confidence intervals ...
```

```r
broom::tidy(mod2a.ci)
```

```
##                         .rownames     X2.5..   X97.5..
## 1         sd_(Intercept)|Subject  3.7716191  6.802803
## 2 cor_Time.c.(Intercept)|Subject  0.6959913  1.000000
## 3               sd_Time.c|Subject  2.3960967  5.098662
## 4                           sigma 13.5566382 15.162105
## 5                     (Intercept)  3.8872922  7.665837
## 6                          Time.c  1.1543336  4.092051
## 7                             SES -0.7654857  2.248586
```

Table of model with nominal predictor

```r
table1 <- table_fun(mod1a)
options(knitr.kable.NA = '')
knitr::kable(table1, caption = "Nominal Model")
```

Table 1: Nominal Model

type	term	estimate	CI
Fixed Parts	(Intercept)	6.07	(3.87, 8.76)
Fixed Parts	Time.c	2.63	(1.36, 3.72)
Fixed Parts	Child.GenderFemale	-0.60	(-4.33, 2.56)
Random Parts	τ_{00}	28.43	(11.94, 46.47)
Random Parts	τ_{11}	14.52	(6.84, 25.84)
Random Parts	τ_{10}	1.00	(0.52, 1.00)
Random Parts	$\hat{\sigma}^2$	207.76	(182.11, 234.48)

Table of model with continuous predictor

```r
table2 <- table_fun(mod2a)
```

Table 2: Continuous Model

type	term	estimate	CI
Fixed Parts	(Intercept)	5.78	(4.20, 7.61)
Fixed Parts	Time.c	2.64	(1.15, 4.25)
Fixed Parts	SES	0.73	(-1.02, 1.79)
Random Parts	τ_{00}	27.79	(12.63, 50.88)
Random Parts	τ_{11}	14.53	(7.00, 26.28)
Random Parts	τ_{10}	1.00	(0.52, 1.00)
Random Parts	$\hat{\sigma}^2$	207.67	(184.17, 229.45)

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

```
mod3 <- lmer(Utterances.with.Letters ~ Time*Child.Gender*SES + (1|Subject), data = children)
summary(mod3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Utterances.with.Letters ~ Time * Child.Gender * SES + (1 | Subject)
##    Data: children
##
## REML criterion at convergence: 5414.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4280 -0.3531 -0.1579  0.0137 10.4621
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Subject  (Intercept)  22.7     4.764
##  Residual             229.3    15.144
## Number of obs: 652, groups:  Subject, 55
##
## Fixed effects:
##                             Estimate Std. Error t value
## (Intercept)                   0.4702     1.7944   0.262
## Time                          3.2063     0.7205   4.450
## Child.GenderFemale            1.3477     2.6002   0.518
## SES                          -1.2529     2.1517  -0.582
## Time:Child.GenderFemale      -1.5853     1.0406  -1.523
## Time:SES                      2.0292     0.8641   2.348
## Child.GenderFemale:SES        2.1404     2.7117   0.789
## Time:Child.GenderFemale:SES  -2.3718     1.0856  -2.185
##
## Correlation of Fixed Effects:
##             (Intr) Time   Chl.GF SES    Tm:C.GF Tm:SES C.GF:S
## Time        -0.733
## Chld.GndrFm -0.690  0.506
## SES         -0.164  0.122  0.113
## Tm:Chld.GnF  0.507 -0.692 -0.732 -0.084
## Time:SES     0.122 -0.171 -0.084 -0.733  0.119
## Chld.GF:SES  0.130 -0.097 -0.033 -0.793  0.026   0.582
## Tm:C.GF:SES -0.097  0.136  0.026  0.584 -0.039  -0.796 -0.733
```

Fixed Effects

Intercept: 0.47; mean of number of utterances with letters for males with mean level SES at Session 1

Time: 3.21; increase in number of utterances with letters every year, for males and when SES at mean level

Child.Gender: 1.35; the difference in the number of utterances with letters between males and females when SES at mean level

SES: -1.25; decrease, for males, in number of utterances for every 1 unit increase in SES at Session 1

Session:Child.Gender: -1.59; the difference between the slopes for males and females when SES at mean level

Time:SES: 2.03; the change in the slope of Time, for males, for every 1 unit increase in SES

Child.Gender:SES: 2.14; the difference between the relationship of SES and the number of utterances with letters for males and females at Session 1

Time:Child.Gender:SES: -2.37; the difference between the interaction of SES and Time for males and females

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

```r
mod4 <- lmer(Utterances.with.Letters ~ Time.c + Utterances + (Time.c|Subject), data = children)
summary(mod4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula:
## Utterances.with.Letters ~ Time.c + Utterances + (Time.c | Subject)
##    Data: children
##
## REML criterion at convergence: 5402.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.4959 -0.2900 -0.1330 -0.0361 10.0466
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept)  22.10    4.701
##           Time.c       13.03    3.610   1.00
##  Residual             207.93   14.420
## Number of obs: 652, groups:  Subject, 55
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 2.4290087  1.4766579   1.645
## Time.c      1.2333047  0.8548480   1.443
## Utterances  0.0022725  0.0008182   2.777
##
## Correlation of Fixed Effects:
##            (Intr) Time.c
## Time.c      0.726
## Utterances -0.818 -0.590
## fit warnings:
## Some predictor variables are on very different scales: consider rescaling
```

Fixed Effects

Intercept: -0.19; mean of number of utterances with letters for children at Session 1

Time: 0.58; increase in number of utterances with letters every year

Utterances: 0.0033; increase in number of utterances for every 1 unit increase number of utterances