

Homework 5

JP

2/5/2020

<https://quantdev.ssri.psu.edu/tutorials/growth-modeling-chapter-3-linear-growth-models>

The following is a basic description of variables you will be using in this assignment. § aid = a unique id number assigned to each adolescent respondent § schoolid = a unique id number assigned to each school in the data set § bmi_w1 = a continuous measure of body mass index (BMI) calculated from respondent height and weight in wave 1 of the survey (collected in 1994-95). § bmi_w2 = a continuous measure of body mass index (BMI) calculated from respondent height and weight in wave 2 of the survey (collected in 1996). § bmi_w3 = a continuous measure of body mass index (BMI) calculated from respondent height and weight in wave 3 of the survey collected in 2001-2002, when most respondents were between ages 18 and 26). § bmi_w4 = a continuous measure of body mass index (BMI) calculated from respondent height and weight in wave 4 of the survey (collected in 2007-2008, when most respondents were between the ages of 24 and 32). § sex = adolescent sex (it is unclear whether this survey item was more closely measuring sex or gender. For our purposes assume it measures gender.) o 1 = male o 2 = female § parent_highestedu = highest educational attainment of either parent/guardian/parent-figure of adolescents in the sample: o 1 = Less than high school (no HS degree) o 2 = Completed high school or equivalent o 3 = Some college (no degree) o 4 = College degree or more § age_w1 = age in years (continuous measure) reported at wave 1 § multpov_w1 = a ratio of the family's income to the poverty threshold for a household of their size, as measured in wave 1. multpov_w1 is therefore a multiple of the poverty line. If multpov_w1 > 0 then their household is above the federal poverty line, if multpov_w1 < 0 then it is below the federal poverty line.

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 3.5.3
```

```
## -- Attaching packages -----
```

```
## v ggplot2 3.2.1      v purrr  0.3.2
## v tibble  2.1.1      v dplyr  0.8.3
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
## Warning: package 'tibble' was built under R version 3.5.3
```

```
## Warning: package 'tidyr' was built under R version 3.5.3
```

```
## Warning: package 'readr' was built under R version 3.5.3
```

```
## Warning: package 'purrr' was built under R version 3.5.3
```

```
## Warning: package 'dplyr' was built under R version 3.5.3
```

```
## Warning: package 'stringr' was built under R version 3.5.2
```

```
## Warning: package 'forcats' was built under R version 3.5.3
```

```
## -- Conflicts -----  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()     masks stats::lag()
```

```
library(readstata13)
```

```
## Warning: package 'readstata13' was built under R version 3.5.3
```

```
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 3.5.3
```

```
## Loading required package: Matrix
```

```
##  
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':  
##  
##     expand, pack, unpack
```

```
library(psych)
```

```
## Warning: package 'psych' was built under R version 3.5.3
```

```
##  
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':  
##  
##     %+%, alpha
```

```
library(optimx)
```

```
## Warning: package 'optimx' was built under R version 3.5.3
```

```
library(lmerTest)
```

```
## Warning: package 'lmerTest' was built under R version 3.5.3
```

```
##  
## Attaching package: 'lmerTest'
```

```
## The following object is masked from 'package:lme4':  
##  
##     lmer
```

```
## The following object is masked from 'package:stats':  
##  
##     step
```

```
library(dfoptim)
```

```
## Warning: package 'dfoptim' was built under R version 3.5.2
```

```
options(max.print = 99999)
```

```
options(scipen = 999)
```

```
getwd()
```

```
## [1] "E:/U0/R Projects/SOC 613/scripts"
```

```
set.seed(232020)
```

```
data <- read.dta13("E:/U0/R Projects/SOC 613/data/AddHealth.dta")
```

```
data <- data %>%  
  dplyr::select(aid,  
                schoolid,  
                bmi_w1,  
                bmi_w2,  
                bmi_w3,  
                bmi_w4,  
                sex,  
                parent_highestedu,  
                age_w1,  
                multpov_w1)
```

```
colnames(data)
```

```
## [1] "aid"          "schoolid"      "bmi_w1"  
## [4] "bmi_w2"      "bmi_w3"        "bmi_w4"  
## [7] "sex"         "parent_highestedu" "age_w1"  
## [10] "multpov_w1"
```

Data manipulation: a. Convert the Add Health data set from wide format to long format, generating new variables as you do—“wave” and “bmi.” Be sure that in your final data set you also have the following variables: adolescent id (aid), school id (schoolid), sex, multpov_w1, parent_highestedu, and age_w1. b. Create variables “female” and “poor_w1” coded as we did in Homework #1. c. When you treat wave as a predictor in these models you would like the intercept to be the baseline measurement that occurred at Wave 1. Create a new variable called “time” that is coded as follows: Time = 0 if Wave 1 Time = 1 if Wave 2 Time = 2 if Wave 3 Time = 3 if Wave 4

```
long <- data %>%  
  gather(key = time, value = bmi_values, c(-1:-2, -7:-10)) %>%  
  mutate(female = recode(sex, "1" = 'male',  
                           "2" = 'female'),  
         poor = case_when(multpov_w1 <= 1 ~ 1,  
                           multpov_w1 > 1 ~ 0),  
         poor = recode(poor, '1' = 'poor',  
                        '0' = 'not_poor'),  
         female = as.factor(female),
```

```

poor = as.factor(poor),
aid = as.numeric(aid),
time = as.factor(time)) %>%
mutate(parent_highestedu = recode(parent_highestedu, '1' = 'less_than_hs',
                                '2' = 'hs_grad',
                                '3' = 'some_college',
                                '4' = 'college_degree'),
sex = recode(sex, '1' = 'male',
              '2' = 'female'),
time = recode(time, 'bmi_w1' = '0',
              'bmi_w2' = '1',
              'bmi_w3' = '2',
              'bmi_w4' = '3')) %>%
mutate(parent_highestedu = relevel(as.factor(parent_highestedu), ref = 'college_degree'),
sex = relevel(as.factor(sex), ref = 'male')) %>%
mutate(bmi_values = as.numeric(bmi_values),
time = as.numeric(time),
age_w1 = as.numeric(age_w1),
aid = as.numeric(aid))

```

Write the following models using proper notation and show all steps (micro model, macro model, combined model). In all cases the model will be two-level structures where observations (level 1) are nested in adolescents (level 2). Ignore clustering of adolescents within schools for the sake of this exercise. The outcome for all models will be “bmi.” Do not provide interpretations at this stage. a. Model 1: Two-level Random Intercept model with time as a continuous fixed effect predictor.

$$BMI_{ti} = \beta_{0i} + \beta_{1i}(time_{ti}) + \epsilon_{0ti}$$

$$\beta_{0i} = \beta_0 + \mu_{0i}$$

$$BMI_{ti} = \beta_0 + \beta_1(time_{ti}) + \mu_{0i}(time_{0ti}) + \epsilon_{0ti}$$

$$Level2 : [\mu_{0i}] \sim N(0, \sigma_{\mu_0}^2)$$

$$Level1 : [\epsilon_{0ti}] \sim N(0, \sigma_{\epsilon_0}^2)$$

- b. Model 2: Two-level Random Intercept model with the following FE predictors: time, female, poor_w1, parent_highestedu, and age_w1. Treat “college degree or more” as the reference category for parent_highestedu.

$$BMI_{ti} = \beta_{0i} + \beta_{1i}(time_{ti}) + \beta_{2i}(female_{ti}) + \beta_{3i}(poor_{ti}) + \beta_{4i}(less_than_hs_{ti}) + \beta_{5i}(highschool_{ti}) + \beta_{6i}(some_college_{ti}) + \beta_{7i}(age_{ti}) +$$

$$\beta_{0i} = \beta_0 + \mu_{0i}$$

$$BMI_{ti} = \beta_0 + \beta_1(time_{ti}) + \beta_2(female_{ti}) + \beta_3(poorn_{ti}) + \beta_4(less_than_hs_{ti}) + \beta_5(highschool_{ti}) + \beta_6(some_college_{ti}) + \beta_7(age_{ti}) + \mu_{0i} +$$

$$Level2 : [\mu_{0i}] \sim N(0, \sigma_{\mu 0}^2)$$

$$Level1 : [\epsilon_{0ti}] \sim N(0, \sigma_{\epsilon 0}^2)$$

- c. Model 3: Two-level Random Slopes model with time as a random slopes variable and the following variables as additional fixed effects: female, poor_w1, parent_highestedu, and age_w1. Treat “college degree or more” as the reference category for parent_highestedu. Constrain the level 2 covariance between random intercepts and random slopes to be = 0.

Micro:

$$BMI_{ti} = \beta_{0i} + \beta_{1i}(time_{ti}) + \beta_{2i}(female_{ti}) + \beta_{3i}(poor_{ti}) + \beta_{4i}(lessthanhs_{ti}) + \beta_{5i}(highschool_{ti}) + \beta_{6i}(somecollege_{ti}) + \beta_{7i}(age_{ti}) +$$

Macro 1:(Intercept)

$$\beta_{0i} = \beta_0 + \mu_{0i}$$

Macro 2: (Slope)

$$\beta_{1i} = \beta_1 + \mu_{1i}$$

Combined:

$$BMI_{ti} = \beta_0 + \beta_1(time_{ti}) + \beta_2(female_{ti}) + \beta_3(poorn_{ti}) + \beta_4(lessthanhs_{ti}) + \beta_5(highschool_{ti}) + \beta_6(somecollege_{ti}) + \beta_7(age_{ti}) + \mu_{0i} +$$

$$Level2 : \begin{bmatrix} \mu_{0i} \\ \mu_{1i} \end{bmatrix} \sim N \left(0, \begin{bmatrix} \sigma_{\mu 0}^2 & 0 \\ 0 & \sigma_{\mu 1}^2 \end{bmatrix} \right)$$

$$Level1 : [\epsilon_{0ti}] \sim N(0, \sigma_{\epsilon 0}^2)$$

- d. Model 4: Same as Model 3, except that now you also allow for an interaction between female gender and time. Include “female” main effects as well as interactions with the time variable.

Micro:

$$BMI_{ti} = \beta_{0i} + \beta_{1i}(time_{ti}) + \beta_{2i}(female_{ti}) + \beta_{3i}(poor_{ti}) + \beta_{4i}(lessthanhs_{ti}) + \beta_{5i}(highschool_{ti}) + \beta_{6i}(somecollege_{ti}) + \beta_{7i}(age_{ti}) +$$

Macro 1:(Intercept)

$$\beta_{0i} = \beta_0 + \mu_{0i}$$

Macro 2: (Slope)

$$\beta_{1i} = \beta_1 + \mu_{1i}$$

Combined:

$$BMI_{ti} = \beta_0 + \beta_1(time_{ti}) + \beta_2(female_{ti}) + \beta_3(poorn_{ti}) + \beta_4(lessthanhs_{ti}) + \beta_5(highschool_{ti}) + \beta_6(somecollege_{ti}) + \beta_7(age_{ti}) + \beta_8(ti$$

$$Level2 : \begin{bmatrix} \mu_{0i} \\ \mu_{1i} \end{bmatrix} \sim N \left(0, \begin{bmatrix} \sigma_{\mu 0}^2 & 0 \\ 0 & \sigma_{\mu 1}^2 \end{bmatrix} \right)$$

$$Level1 : [\epsilon_{0ti}] \sim N(0, \sigma_{\epsilon 0}^2)$$

Answer the following questions about the models you wrote in Question 2: a. Model 1 Questions: i. What are the interpretations for the two variance (random effect) parameters in this model?

Answer: μ_{0i} is the variation found between individuals across time. ϵ_{0ti} is the variation within each individual across time.

- ii. What does the beta parameter “time” explain in this model? In other words, is the “change over time” a change within-individuals or across all individuals?

Answer: The fixed effect of time explains the overall association between time and BMI. The change in BMI over time is across all individuals.

- b. Model 2 Questions:
- c. How do we expect that adding the level 2 variables will affect each of the two variance parameters of the model?

Answer: The addition of level 2 variables will reduce variation across all individuals, so the μ_{0i} should be lower as these fixed effects should affect this value. Within-individual differences should not change much.

- ii. What does including a main effect for “female” imply about how we think female versus male gender affects BMI trajectories?

Answer: The main effect for female implies that the trajectory between males and females should be different. Such as across all observations, females may have lower BMI than males.

- c. Model 3 Questions:
- d. What does the beta parameter for “time” represent in this model?

Answer: The fixed effect of time is the overall association between time and BMI across all individuals and observations.

- ii. What does it mean to allow “time” to be a random variable in this model?

Answer: By allowing time to be a random variable in this model, we can examine the variation between individuals in the slope of the association between time and BMI. We can examine as time moves forward, what their BMI will look like and see if there is variation between individuals.

In other words, what does this allow for?

Answer: This allows to examine if individuals’ BMI varies over time, rather than an overall examination of all individuals’ BMI over time in the fixed effect. We can then examine if there are certain characteristics of some individuals that have different trajectories of BMI over time.

- d. Model 4 Questions:
- e. What does it mean that we are allowing for “female” gender to interact with “time” in this model? In other words, how does it differ from what we did in Question 2?

Answer: Since we are allowing for an interaction between the fixed effects of female and time, we are looking at the association between sex and BMI across all observations. This differs slightly from question 2 because we are interested in examining if the overall difference between the two sexes are different as time goes on in this example, while question 2, we were interested in if there were differences across all 4 waves collectively.

Fit all four models outlined in Question 2 using STATA. When fitting Models 1 and 2 also calculate the ICC.

```
long %>%  
  colnames()
```

```
## [1] "aid"           "schoolid"       "sex"
## [4] "parent_highestedu" "age_w1"         "multpov_w1"
## [7] "time"           "bmi_values"     "female"
## [10] "poor"
```

```
time_only <- lmer(bmi_values ~ time + 1 +
                  (1 | aid), data = long, REML = FALSE,
                  control = lmerControl(optimizer = 'Nelder_Mead'))
summary(time_only)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use
## Satterthwaite's method [lmerModLmerTest]
## Formula: bmi_values ~ time + 1 + (1 | aid)
## Data: long
## Control: lmerControl(optimizer = "Nelder_Mead")
##
##      AIC      BIC    logLik deviance df.resid
## 116642.0 116673.8 -58317.0 116634.0    20768
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.6044 -0.5127 -0.0583  0.4347  7.4933
##
## Random effects:
## Groups   Name      Variance Std.Dev.
## aid      (Intercept) 22.554   4.749
## Residual                7.896   2.810
## Number of obs: 20772, groups: aid, 6491
##
## Fixed effects:
##              Estimate Std. Error      df t value
## (Intercept)  20.16816    0.07471 12254.32594   269.9
## time         1.95315    0.01747 14678.84684   111.8
##              Pr(>|t|)
## (Intercept) <0.0000000000000002 ***
## time        <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## time -0.549
```

```
intercept <- lmer(bmi_values ~ time + sex + poor + parent_highestedu + age_w1 +
                  1 +
                  (1 | aid), data = long, REML = FALSE,
                  control = lmerControl(optimizer = 'Nelder_Mead'))
summary(intercept)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use
## Satterthwaite's method [lmerModLmerTest]
## Formula: bmi_values ~ time + sex + poor + parent_highestedu + age_w1 +
## 1 + (1 | aid)
```

```

## Data: long
## Control: lmerControl(optimizer = "Nelder_Mead")
##
##      AIC      BIC    logLik deviance df.resid
## 89403.7 89480.5 -44691.9 89383.7    15962
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.6491 -0.5082 -0.0571  0.4373  7.4998
##
## Random effects:
## Groups Name Variance Std.Dev.
## aid (Intercept) 21.970 4.687
## Residual 7.837 2.799
## Number of obs: 15972, groups: aid, 4909
##
## Fixed effects:
##
## Estimate Std. Error df t value
## (Intercept) 14.28288 0.64301 5002.24258 22.213
## time 1.95032 0.01986 11324.98266 98.224
## sexfemale -0.14900 0.14191 4898.74848 -1.050
## poorpoor 0.28347 0.19458 4924.92832 1.457
## parent_highesteduhs_grad 1.04856 0.18873 4892.47763 5.556
## parent_highesteduleless_than_hs 1.45622 0.28703 4946.31486 5.073
## parent_highestedusome_college 0.75762 0.17525 4884.01328 4.323
## age_w1 0.34039 0.04066 4967.05080 8.371
##
## Pr(>|t|)
## (Intercept) < 0.0000000000000002 ***
## time < 0.0000000000000002 ***
## sexfemale 0.294
## poorpoor 0.145
## parent_highesteduhs_grad 0.0000000291 ***
## parent_highesteduleless_than_hs 0.0000004051 ***
## parent_highestedusome_college 0.0000156935 ***
## age_w1 < 0.0000000000000002 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
## (Intr) time sexfml poorpr prnt_hghstdh_ prn___
## time -0.076
## sexfemale -0.153 -0.009
## poorpoor -0.036 0.004 0.010
## prnt_hghstdh_ -0.090 0.001 -0.012 -0.212
## prnt_hghs__ -0.007 0.002 -0.003 -0.342 0.328
## prnt_hghstds_ -0.110 0.002 0.017 -0.101 0.438 0.309
## age_w1 -0.974 0.004 0.041 0.015 -0.019 -0.063
## prnt_hghstds_
## time
## sexfemale
## poorpoor
## prnt_hghstdh_
## prnt_hghs__
## prnt_hghstds_

```



```
## age_w1          -0.013
```

<https://stats.stackexchange.com/questions/13166/rs-lmer-cheat-sheet/13173#13173>

```
str(long)
```

```
## 'data.frame':   26016 obs. of  10 variables:
## $ aid          : num  57100270 57101310 57103171 57103869 57104553 ...
## $ schoolid     : num  178 108 115 167 142 278 278 224 256 185 ...
## $ sex          : Factor w/ 2 levels "male","female": 2 2 1 1 2 1 1 1 1 1 ...
## $ parent_highestedu: Factor w/ 4 levels "college_degree",...: 1 4 2 2 1 1 2 3 2 1 ...
## $ age_w1       : num   17 18 15 18 19 13 11 14 13 14 ...
## $ multpov_w1   : num   3.056 NA 3.75 0.45 0.105 ...
## $ time         : num   1 1 1 1 1 1 1 1 1 1 ...
## $ bmi_values   : num   40.3 20.6 NA 24.1 31.4 ...
## $ female       : Factor w/ 2 levels "female","male": 1 1 2 2 1 2 2 2 2 2 ...
## $ poor         : Factor w/ 2 levels "not_poor","poor": 1 NA 1 2 2 1 1 2 2 1 ...
```

```
long %>%
  group_by(aid) %>%
  summarize(n = n()) %>%
  ungroup() %>%
  count(aid) %>%
  summarize(individual = sum(n))
```

```
## # A tibble: 1 x 1
##   individual
##       <int>
## 1         6504
```

```
test <- lmer(bmi_values ~ time + sex + poor + parent_highestedu + age_w1 +
             (time || aid), data = long, REML = FALSE,
             control = lmerControl(optimizer = 'Nelder_Mead'))
summary(test)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use
## Satterthwaite's method [lmerModLmerTest]
## Formula: bmi_values ~ time + sex + poor + parent_highestedu + age_w1 +
##         (time || aid)
## Data: long
## Control: lmerControl(optimizer = "Nelder_Mead")
##
##      AIC      BIC    logLik deviance df.resid
## 87128.0 87212.5 -43553.0 87106.0    15961
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -6.0292 -0.4189 -0.0227  0.3805  6.3732
##
## Random effects:
## Groups   Name                Variance Std.Dev.
## aid      (Intercept) 13.317    3.649
```

```

## aid.1    time      1.629    1.276
## Residual      5.028    2.242
## Number of obs: 15972, groups: aid, 4909
##
## Fixed effects:
##
##              Estimate Std. Error      df t value
## (Intercept)    12.72252    0.56476 4886.32939  22.527
## time           1.93798    0.02506 5596.60556   77.342
## sexfemale     -0.34438    0.12466 4796.34370   -2.763
## poorpoor       0.16540    0.17099 4819.53092    0.967
## parent_highesteduhs_grad 0.66707    0.16590 4800.39440    4.021
## parent_highesteduleless_than_hs 1.21346    0.25270 4864.61003    4.802
## parent_highestedusome_college 0.49280    0.15386 4775.91358    3.203
## age_w1         0.46216    0.03578 4889.33824   12.916
##
##              Pr(>|t|)
## (Intercept)    < 0.0000000000000002 ***
## time           < 0.0000000000000002 ***
## sexfemale              0.00576 **
## poorpoor              0.33343
## parent_highesteduhs_grad 0.00005889 ***
## parent_highesteduleless_than_hs 0.00000162 ***
## parent_highestedusome_college 0.00137 **
## age_w1           < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) time  sexfml poorpr prnt_hghstdh_ prn___
## time      -0.044
## sexfemale  -0.153 -0.003
## poorpoor   -0.034 -0.003  0.009
## prnt_hghstdh_ -0.090 -0.003 -0.010 -0.212
## prnt_hghs__  -0.006 -0.005 -0.002 -0.342  0.327
## prnt_hghstds_ -0.109  0.000  0.017 -0.101  0.438      0.309
## age_w1      -0.975 -0.002  0.041  0.014 -0.018      -0.063
##      prnt_hghstds_
## time
## sexfemale
## poorpoor
## prnt_hghstdh_
## prnt_hghs__
## prnt_hghstds_
## age_w1      -0.013

```

```
# summary(allFit(test))
```

```

slope_zero_vcov <- lmer(bmi_values ~ time + sex + poor + parent_highestedu + age_w1 +
                        (time || aid), data = long, REML = FALSE,
                        control = lmerControl(optimizer = 'Nelder-Mead'))

summary(slope_zero_vcov)

```

```
## Linear mixed model fit by maximum likelihood . t-tests use
```

```

## Satterthwaite's method [lmerModLmerTest]
## Formula: bmi_values ~ time + sex + poor + parent_highestedu + age_w1 +
## (time || aid)
## Data: long
## Control: lmerControl(optimizer = "Nelder_Mead")
##
##      AIC      BIC    logLik deviance df.resid
## 87128.0 87212.5 -43553.0 87106.0    15961
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -6.0292 -0.4189 -0.0227  0.3805  6.3732
##
## Random effects:
## Groups   Name                Variance Std.Dev.
## aid      (Intercept)    13.317    3.649
## aid.1    time              1.629    1.276
## Residual                    5.028    2.242
## Number of obs: 15972, groups: aid, 4909
##
## Fixed effects:
##
##              Estimate Std. Error      df t value
## (Intercept)    12.72252    0.56476 4886.32939  22.527
## time            1.93798    0.02506 5596.60556  77.342
## sexfemale      -0.34438    0.12466 4796.34370  -2.763
## poorpoor        0.16540    0.17099 4819.53092   0.967
## parent_highesteduhs_grad  0.66707    0.16590 4800.39440   4.021
## parent_highesteduleless_than_hs 1.21346    0.25270 4864.61003   4.802
## parent_highestedusome_college  0.49280    0.15386 4775.91358   3.203
## age_w1          0.46216    0.03578 4889.33824  12.916
##
##              Pr(>|t|)
## (Intercept)    < 0.0000000000000002 ***
## time           < 0.0000000000000002 ***
## sexfemale      0.00576 **
## poorpoor       0.33343
## parent_highesteduhs_grad  0.00005889 ***
## parent_highesteduleless_than_hs 0.00000162 ***
## parent_highestedusome_college  0.00137 **
## age_w1         < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr) time  sexfml poorpr prnt_hghstdh_ prn___
## time          -0.044
## sexfemale      -0.153 -0.003
## poorpoor       -0.034 -0.003  0.009
## prnt_hghstdh_  -0.090 -0.003 -0.010 -0.212
## prnt_hghs__    -0.006 -0.005 -0.002 -0.342  0.327
## prnt_hghstds_  -0.109  0.000  0.017 -0.101  0.438    0.309
## age_w1         -0.975 -0.002  0.041  0.014 -0.018    -0.063
##
##      prnt_hghstds_
## time
## sexfemale

```

```
## poorpoor
## prnt_hghstdh_
## prnt_hghs__
## prnt_hghstds_
## age_w1          -0.013

slope_interaction <- lmer(bmi_values ~ time*sex + poor + parent_highestedu + age_w1 +
  (time || aid), data = long, REML = FALSE,
  control = lmerControl(optimizer = 'Nelder_Mead'))

summary(slope_interaction)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use
## Satterthwaite's method [lmerModLmerTest]
## Formula: bmi_values ~ time * sex + poor + parent_highestedu + age_w1 +
## (time || aid)
## Data: long
## Control: lmerControl(optimizer = "Nelder_Mead")
##
##      AIC      BIC    logLik deviance df.resid
## 87120.6 87212.8 -43548.3 87096.6    15960
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -6.0229 -0.4174 -0.0218  0.3822  6.3960
##
## Random effects:
## Groups   Name                Variance Std.Dev.
## aid      (Intercept)    13.312     3.649
## aid.1    time              1.626     1.275
## Residual                    5.026     2.242
## Number of obs: 15972, groups: aid, 4909
##
## Fixed effects:
##                                Estimate Std. Error      df t value
## (Intercept)                   12.80762    0.56530 4909.96397  22.656
## time                          1.85838    0.03605 5630.75665  51.550
## sexfemale                     -0.50487    0.13517 6182.93370  -3.735
## poorpoor                      0.16543    0.17095 4820.63786   0.968
## parent_highesteduhs_grad      0.66690    0.16587 4801.51046   4.021
## parent_highesteduleless_than_hs 1.21395    0.25264 4865.71593   4.805
## parent_highestedusome_college  0.49281    0.15382 4777.03745   3.204
## age_w1                       0.46199    0.03577 4890.45763  12.914
## time:sexfemale                0.15387    0.05012 5603.89752   3.070
##
##                                Pr(>|t|)
## (Intercept) < 0.0000000000000002 ***
## time < 0.0000000000000002 ***
## sexfemale 0.000189 ***
## poorpoor 0.333231
## parent_highesteduhs_grad 0.00005892 ***
## parent_highesteduleless_than_hs 0.00000159 ***
## parent_highestedusome_college 0.001365 **
## age_w1 < 0.0000000000000002 ***
## time:sexfemale 0.002148 **
```

```

## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) time   sexfml  poorpr  prnt_hghstdh_  prn___
## time          -0.065
## sexfemale      -0.160   0.277
## poorpoor       -0.034 -0.002   0.008
## prnt_hghstdh_  -0.090 -0.001 -0.009 -0.212
## prnt_hghs__    -0.006 -0.004 -0.002 -0.342   0.327
## prnt_hghstds_  -0.109  0.000  0.016 -0.101   0.438           0.309
## age_w1         -0.974 -0.001  0.038  0.014 -0.018           -0.063
## time:sexfml    0.049 -0.719 -0.387  0.000 -0.001           0.000
##              prnt_hghstds_  age_w1
## time
## sexfemale
## poorpoor
## prnt_hghstdh_
## prnt_hghs__
## prnt_hghstds_
## age_w1        -0.013
## time:sexfml   0.000          -0.001

```

Answer the following questions using your results from STATA: a. Across the four models, which beta parameters are statistically significant at 0.05, and what does it mean that they are significant (consider also the direction of any effects and what that indicates)?

Answer: Model 1: Across all individuals and observations, the average BMI increased. As time went on, BMI increased for all individuals.

Model 2: Adjusting for time, sex, poor, parent's education, and age, the average individual's BMI increased. As time went on, the overall average individual's BMI increased. Individuals with parents that had any education lower than a college degree were found to have a higher BMI across all observations. The older an individual was, the higher the overall BMI increased.

Model 3: Adjusting for time, sex, poor, parent's education, and age, the average individual's BMI increased. As time went on, the overall average individual's BMI increased. Across all observations, females had a lower BMI than males. Individuals with parents that had any education lower than a college degree were found to have a higher BMI across all observations. The older an individual was, the higher the overall BMI increased.

Model 4: Adjusting for time, sex, poor, parent's education, and age, the average individual's BMI increased. As time went on, the overall average individual's BMI increased. Across all observations, females had a lower BMI than males. Individuals with parents that had any education lower than a college degree were found to have a higher BMI across all observations. The older an individual was, the higher the overall BMI increased. As time went on, females had an overall larger increase in BMI than males.

- b. How do the variance parameters change across the four models (i.e., Model 1 to Model 2, Model 2 to Model 3, Model 3 to Model 4)? Why do you expect that the changes at each step occurred as they did?

Answer: The variation between individuals in model 1 was 22.55 and decreased to 21.97. This change was because the additional variables accounted for additional variance in explaining differences in BMI between individuals. These variables did not appear to contribute much to a difference within individuals as these values were roughly the same (7.90 to 7.84). The differences between model 2 and model 3 showed that

the differences in BMI between individuals changed when accounting for the differences in slopes of time and BMI. The differences between individuals' BMI changed from 21.97 to 13.32, which could be due to the incorporation of the slopes changing between time and BMI for individuals. Now accounting for the differing slopes of time and BMI between individuals, there was less variation within individuals also (7.84 to 5.03). Lastly, with the inclusion of the fixed effect interaction between time and sex on BMI, there didn't seem to be much of a difference in variation of random effects. This could be because the interaction was focusing on time and females across all observations.

- c. For the two Random Intercept models (Models 1 and 2) what are the ICCs, and what do these values mean? Use the variance partition coefficient interpretation.

Answer: The ICC for the model only including time was .64, while the model that adjusted for additional variables (e.g., age, sex, parent education, poverty) had an ICC of .62. For the first model, 64% of the variation in BMI is attributable to differences between individuals. While adjusting for the additional variables, 62% of the variation in BMI was attributable to differences between individuals.

```
performance::icc(time_only)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.741
##      Conditional ICC: 0.635
```

```
performance::icc(intercept)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.737
##      Conditional ICC: 0.619
```