Homework 8

JP

3/3/2020

# Homework Assignment #8

In this assignment you will again be working with a subset of the publicly available version of the data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). This data can be found on the course website in Canvas under the “Files” tab. Look for data file “AddHealth.dta.” Use the STATA commands handout (also available in the Assignments tab) to help youcomplete this assignment. On the due date please hand in a hardcopy of your STATA outputand answers to all questions in this assignment. You can either hand in a document with youranswers (e.g., Word doc) with STATA output attached or you can incorporate answers to the questions directly into the STATA output in the form of comments. I prefer typed assignments inTimes New Roman size 12 or Arial size 11.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

library(tidyverse)

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

## -- Attaching packages ------------------------ tidyverse 1.3.0 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 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 --------------------------- tidyverse\_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(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(nlme)

##   
## Attaching package: 'nlme'

## The following object is masked from 'package:dplyr':  
##   
## collapse

library(robustlmm)

## Warning: package 'robustlmm' was built under R version 3.5.3

## Loading required package: 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

##   
## Attaching package: 'lme4'

## The following object is masked from 'package:nlme':  
##   
## lmList

library(lme4)  
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

options(max.print = 99999)  
options(scipen = 999)  
  
getwd()

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

set.seed(342020)  
  
health <- read.dta13("E:/UO/R Projects/SOC 613/data/AddHealth.dta")  
  
health\_sub <- health %>%   
 dplyr::select(aid,  
 schoolid,  
 bmi\_w1,  
 bmi\_w2,  
 bmi\_w3,  
 bmi\_w4,  
 sex,  
 parent\_highestedu,  
 age\_w1,  
 multpov\_w1)  
  
health\_sub %>%   
colnames()

## [1] "aid" "schoolid" "bmi\_w1"   
## [4] "bmi\_w2" "bmi\_w3" "bmi\_w4"   
## [7] "sex" "parent\_highestedu" "age\_w1"   
## [10] "multpov\_w1"

## Tasks:

In this assignment you will build on the two-level (panel) models you fit in Homework #5 tocreate three-level models in the Add Health data. Level 1 will be time (or wave) of observation,Level 2 will be adolescents (aid) and Level 3 will be school attended in wave 1 (schoolid).

1. Data manipulation – same as in Homework 5:
2. Convert the Add Health data set from wide format to long format, generating newvariables as you do—“wave” and “bmi.” Be sure that in your final data set youalso have the following variables: adolescent id (aid), school id (schoolid), sex,multpov\_w1, parent\_highestedu, and age\_w1.

long <- health\_sub %>%   
 gather(key = wave, value = bmi\_values, c(-1:-2, -7:-10))

1. Create a variable for “female” gender.

long <- long %>%   
 mutate(sex = recode(sex, '1' = 'male',  
 '2' = 'female'))

1. When you treat wave as a predictor in these models you would like the interceptto be the baseline measurement that occurred at Wave 1. Create a new variable called “time” that is coded as follows:

Time = 0 = Wave 1 Time = 1 = Wave 2 Time = 2 = Wave 3 Time = 3 = Wave 4

long <- long %>%   
 mutate(time = recode(wave, 'bmi\_w1' = '0',  
 'bmi\_w2' = '1',  
 'bmi\_w3' = '2',  
 'bmi\_w4' = '3'))  
  
str(long)

## 'data.frame': 26016 obs. of 9 variables:  
## $ aid : chr "57100270" "57101310" "57103171" "57103869" ...  
## $ schoolid : num 178 108 115 167 142 278 278 224 256 185 ...  
## $ sex : chr "female" "female" "male" "male" ...  
## $ parent\_highestedu: num 4 3 2 2 4 4 2 1 2 4 ...  
## $ 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 ...  
## $ wave : chr "bmi\_w1" "bmi\_w1" "bmi\_w1" "bmi\_w1" ...  
## $ bmi\_values : num 40.3 20.6 NA 24.1 31.4 ...  
## $ time : chr "0" "0" "0" "0" ...

long <- long %>%   
 mutate(time = as.numeric(time))

1. Write the following models using proper notation and show all steps (level 1 model, level 2 model(s), level 3 model(s), combined model). In all cases except for Model 1, the models will have three-level structures where observations (level 1) are nested in adolescents (level 2) who are in turn nested in schools (level 3). The outcome for all models will be “bmi.” You may assume that any covariances between random variables are = 0. Do not provide interpretations at this stage.

\*\*\* These models will build on each other, so be careful not to “loose” predictors you are asked to include in later stages.

1. Model 1: Two-level Linear Random Intercept model with NO fixed effect predictors (null model).
2. Model 2: Three-level Linear Random Intercept model with NO fixed effect predictors (null model).
3. Model 3: Model 2 + add the following fixed effect predictors: time, female gender, age at wave 1, and parent education level. In this model be careful to associated predictors with the appropriate level, and treat “college degree or more” as the reference level for parent education.
4. Model 4: Model 3 + treat time as a random variable at both levels 2 and 3 (this is now a Three-level Linear Random Slopes model).
5. Model 5: Model 4 + treat “female” as a random coefficient at level 3.
6. Fit all models outlined in Question 2 using STATA. Store the results as you fit the models using “estimates store” commands. When fitting Models 1, 2, and 3 also calculate the ICC for each contextual level (level 2 and 3). You may request that STATA do these calculations for you.

colnames(long)

## [1] "aid" "schoolid" "sex"   
## [4] "parent\_highestedu" "age\_w1" "multpov\_w1"   
## [7] "wave" "bmi\_values" "time"

model1 <- lmer(bmi\_values ~ 1 + (1 | aid), data = long, REML = FALSE,  
 control = lmerControl(optimizer = 'Nelder\_Mead'))  
summary(model1)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula: bmi\_values ~ 1 + (1 | aid)  
## Data: long  
## Control: lmerControl(optimizer = "Nelder\_Mead")  
##   
## AIC BIC logLik deviance df.resid   
## 125749.8 125773.6 -62871.9 125743.8 20769   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.8492 -0.5445 -0.1519 0.4223 6.4298   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## aid (Intercept) 21.39 4.625   
## Residual 14.64 3.826   
## Number of obs: 20772, groups: aid, 6491  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 24.7644 0.0638 6453.5236 388.2 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

model2 <- lmer(bmi\_values ~ 1 + (1 | schoolid) + (1 | schoolid:aid), data = long, REML = FALSE,  
 control = lmerControl(optimizer = 'Nelder\_Mead'))  
summary(model2)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula: bmi\_values ~ 1 + (1 | schoolid) + (1 | schoolid:aid)  
## Data: long  
## Control: lmerControl(optimizer = "Nelder\_Mead")  
##   
## AIC BIC logLik deviance df.resid   
## 125612.8 125644.6 -62802.4 125604.8 20768   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -3.8543 -0.5428 -0.1514 0.4217 6.4318   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## schoolid:aid (Intercept) 20.25 4.500   
## schoolid (Intercept) 1.09 1.044   
## Residual 14.64 3.826   
## Number of obs: 20772, groups: schoolid:aid, 6491; schoolid, 132  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 24.6739 0.1125 132.9864 219.3 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

long <- long %>%   
 mutate(parent\_highestedu = recode(parent\_highestedu, '1' = 'less\_than\_hs',  
 '2' = 'hs\_grad',  
 '3' = 'some\_college',  
 '4' = 'college\_degree'),  
 parent\_highestedu = relevel(as.factor(parent\_highestedu), ref = 'college\_degree'),  
 sex = relevel(as.factor(sex), ref = 'male'),  
 bmi\_values = as.numeric(bmi\_values),  
 time = as.numeric(time),  
 age\_w1 = as.numeric(age\_w1),  
 aid = as.numeric(aid))  
  
model3 <- lmer(bmi\_values ~ time + sex + age\_w1 + parent\_highestedu + (1 | schoolid) + (1 | schoolid:aid), data = long, REML = FALSE,  
 control = lmerControl(optimizer = 'Nelder\_Mead'))  
summary(model3)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula:   
## bmi\_values ~ time + sex + age\_w1 + parent\_highestedu + (1 | schoolid) +   
## (1 | schoolid:aid)  
## Data: long  
## Control: lmerControl(optimizer = "Nelder\_Mead")  
##   
## AIC BIC logLik deviance df.resid   
## 114903.3 114982.6 -57441.6 114883.3 20510   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -4.5943 -0.5111 -0.0571 0.4366 7.4944   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## schoolid:aid (Intercept) 21.1057 4.5941   
## schoolid (Intercept) 0.6563 0.8101   
## Residual 7.8670 2.8048   
## Number of obs: 20520, groups: schoolid:aid, 6403; schoolid, 132  
##   
## Fixed effects:  
## Estimate Std. Error df t value  
## (Intercept) 16.87983 0.62489 1737.83659 27.012  
## time 1.95412 0.01755 14508.57979 111.372  
## sexfemale -0.03751 0.12301 6374.26430 -0.305  
## age\_w1 0.30136 0.03977 1599.74559 7.578  
## parent\_highesteduhs\_grad 0.87021 0.16506 5618.87832 5.272  
## parent\_highesteduless\_than\_hs 1.26091 0.23080 5426.70563 5.463  
## parent\_highestedusome\_college 0.68468 0.15857 6286.45258 4.318  
## Pr(>|t|)   
## (Intercept) < 0.0000000000000002 \*\*\*  
## time < 0.0000000000000002 \*\*\*  
## sexfemale 0.76   
## age\_w1 0.0000000000000589 \*\*\*  
## parent\_highesteduhs\_grad 0.0000001399254271 \*\*\*  
## parent\_highesteduless\_than\_hs 0.0000000488394847 \*\*\*  
## parent\_highestedusome\_college 0.0000159918774669 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) time sexfml age\_w1 prnt\_hghstdh\_ prn\_\_\_  
## time -0.040   
## sexfemale -0.145 -0.010   
## age\_w1 -0.972 0.003 0.044   
## prnt\_hghstdh\_ -0.070 0.002 -0.001 -0.048   
## prnt\_hghs\_\_ -0.021 0.003 0.005 -0.067 0.346   
## prnt\_hghstds\_ -0.093 0.001 0.015 -0.026 0.461 0.331

model4 <- lmer(bmi\_values ~ time + sex + age\_w1 + parent\_highestedu + (time || schoolid) + (time || schoolid:aid), data = long, REML = FALSE,  
 control = lmerControl(optimizer = 'Nelder\_Mead'))  
summary(model4)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula: bmi\_values ~ time + sex + age\_w1 + parent\_highestedu + (time ||   
## schoolid) + (time || schoolid:aid)  
## Data: long  
## Control: lmerControl(optimizer = "Nelder\_Mead")  
##   
## AIC BIC logLik deviance df.resid   
## 112144.3 112239.5 -56060.2 112120.3 20508   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -6.2103 -0.3965 -0.0107 0.3685 6.4650   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## schoolid.aid time 1.86984 1.3674   
## schoolid.aid.1 (Intercept) 16.00983 4.0012   
## schoolid time 0.06628 0.2574   
## schoolid.1 (Intercept) 0.42572 0.6525   
## Residual 4.76601 2.1831   
## Number of obs: 20520, groups: schoolid:aid, 6403; schoolid, 132  
##   
## Fixed effects:  
## Estimate Std. Error df t value  
## (Intercept) 15.71349 0.55131 1587.29793 28.502  
## time 1.95278 0.03274 136.53572 59.651  
## sexfemale -0.17775 0.10993 6200.36420 -1.617  
## age\_w1 0.38829 0.03507 1432.38273 11.072  
## parent\_highesteduhs\_grad 0.62376 0.14705 5261.25852 4.242  
## parent\_highesteduless\_than\_hs 1.08445 0.20547 5047.82057 5.278  
## parent\_highestedusome\_college 0.52956 0.14152 6068.61194 3.742  
## Pr(>|t|)   
## (Intercept) < 0.0000000000000002 \*\*\*  
## time < 0.0000000000000002 \*\*\*  
## sexfemale 0.105949   
## age\_w1 < 0.0000000000000002 \*\*\*  
## parent\_highesteduhs\_grad 0.000022537 \*\*\*  
## parent\_highesteduless\_than\_hs 0.000000136 \*\*\*  
## parent\_highestedusome\_college 0.000184 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) time sexfml age\_w1 prnt\_hghstdh\_ prn\_\_\_  
## time -0.024   
## sexfemale -0.146 -0.004   
## age\_w1 -0.973 0.009 0.044   
## prnt\_hghstdh\_ -0.073 -0.002 0.001 -0.046   
## prnt\_hghs\_\_ -0.022 -0.002 0.005 -0.067 0.343   
## prnt\_hghstds\_ -0.096 -0.001 0.016 -0.024 0.460 0.330

sex\_dum <- dummy.code(long$sex)  
  
long <- data.frame(long, sex\_dum)  
  
  
model5 <- lmer(bmi\_values ~ time + female + age\_w1 + parent\_highestedu + (time || schoolid) + (time + female || schoolid:aid), data = long, REML = FALSE,  
 control = lmerControl(optimizer = 'Nelder\_Mead'))  
  
summary(model5)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula:   
## bmi\_values ~ time + female + age\_w1 + parent\_highestedu + (time ||   
## schoolid) + (time + female || schoolid:aid)  
## Data: long  
## Control: lmerControl(optimizer = "Nelder\_Mead")  
##   
## AIC BIC logLik deviance df.resid   
## 112134.0 112237.1 -56054.0 112108.0 20507   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -6.2013 -0.3976 -0.0113 0.3681 6.4604   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## schoolid.aid female 2.4239 1.5569   
## schoolid.aid.1 time 1.8578 1.3630   
## schoolid.aid.2 (Intercept) 14.7910 3.8459   
## schoolid time 0.0662 0.2573   
## schoolid.1 (Intercept) 0.4203 0.6483   
## Residual 4.7731 2.1847   
## Number of obs: 20520, groups: schoolid:aid, 6403; schoolid, 132  
##   
## Fixed effects:  
## Estimate Std. Error df t value  
## (Intercept) 15.67938 0.54967 1581.27153 28.525  
## time 1.95248 0.03270 136.58741 59.710  
## female -0.17467 0.10980 6206.20542 -1.591  
## age\_w1 0.39117 0.03499 1430.11612 11.180  
## parent\_highesteduhs\_grad 0.61765 0.14686 5250.12260 4.206  
## parent\_highesteduless\_than\_hs 1.02861 0.20509 5021.25467 5.016  
## parent\_highestedusome\_college 0.51605 0.14126 6054.65427 3.653  
## Pr(>|t|)   
## (Intercept) < 0.0000000000000002 \*\*\*  
## time < 0.0000000000000002 \*\*\*  
## female 0.111727   
## age\_w1 < 0.0000000000000002 \*\*\*  
## parent\_highesteduhs\_grad 0.000026462 \*\*\*  
## parent\_highesteduless\_than\_hs 0.000000547 \*\*\*  
## parent\_highestedusome\_college 0.000261 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) time female age\_w1 prnt\_hghstdh\_ prn\_\_\_  
## time -0.024   
## female -0.140 -0.004   
## age\_w1 -0.974 0.009 0.043   
## prnt\_hghstdh\_ -0.073 -0.002 0.001 -0.046   
## prnt\_hghs\_\_ -0.021 -0.002 0.005 -0.067 0.343   
## prnt\_hghstds\_ -0.096 -0.001 0.016 -0.024 0.460 0.330

1. Using a series of likelihood ratio tests, compare each model to the previous model (Model 2 vs Model 1, Model 3 vs Model 2, and so on). Of these models, which model represents the best fit to the data? Note: in one of these tests you may get an error message saying that “observations differ.” Request the test anyway, and assume that it passes.

anova(model2, model1)

## Data: long  
## Models:  
## model1: bmi\_values ~ 1 + (1 | aid)  
## model2: bmi\_values ~ 1 + (1 | schoolid) + (1 | schoolid:aid)  
## Df AIC BIC logLik deviance Chisq Chi Df  
## model1 3 125750 125774 -62872 125744   
## model2 4 125613 125645 -62802 125605 138.98 1  
## Pr(>Chisq)   
## model1   
## model2 < 0.00000000000000022 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# can't test the differences between models 2 and 3.  
# anova(model3, model2)  
# According to other fit indices (AIC), model 3 (AIC = 114,903.3) fits better than model 2 (AIC = 125,612.8)  
  
anova(model4, model3)

## Data: long  
## Models:  
## model3: bmi\_values ~ time + sex + age\_w1 + parent\_highestedu + (1 | schoolid) +   
## model3: (1 | schoolid:aid)  
## model4: bmi\_values ~ time + sex + age\_w1 + parent\_highestedu + (time ||   
## model4: schoolid) + (time || schoolid:aid)  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)  
## model3 10 114903 114983 -57442 114883   
## model4 12 112144 112239 -56060 112120 2763 2 < 0.00000000000000022  
##   
## model3   
## model4 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(model5, model4)

## Data: long  
## Models:  
## model4: bmi\_values ~ time + sex + age\_w1 + parent\_highestedu + (time ||   
## model4: schoolid) + (time || schoolid:aid)  
## model5: bmi\_values ~ time + female + age\_w1 + parent\_highestedu + (time ||   
## model5: schoolid) + (time + female || schoolid:aid)  
## Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)   
## model4 12 112144 112239 -56060 112120   
## model5 13 112134 112237 -56054 112108 12.355 1 0.0004399 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

1. Using the results you obtain from STATA, and doing calculations when necessary, complete following tables:
2. Variance parameters associated with each level in each model. Model 1 will not have a “school-level” variance parameter.

as.data.frame(VarCorr(model1))

## grp var1 var2 vcov sdcor  
## 1 aid (Intercept) <NA> 21.39042 4.624978  
## 2 Residual <NA> <NA> 14.63868 3.826053

as.data.frame(VarCorr(model2))

## grp var1 var2 vcov sdcor  
## 1 schoolid:aid (Intercept) <NA> 20.249555 4.499951  
## 2 schoolid (Intercept) <NA> 1.089669 1.043872  
## 3 Residual <NA> <NA> 14.641308 3.826396

as.data.frame(VarCorr(model3))

## grp var1 var2 vcov sdcor  
## 1 schoolid:aid (Intercept) <NA> 21.1056827 4.594092  
## 2 schoolid (Intercept) <NA> 0.6563285 0.810141  
## 3 Residual <NA> <NA> 7.8670330 2.804823

# model 1  
sum(insight::get\_variance\_intercept(model1)) /  
(sum(insight::get\_variance\_intercept(model1)) + insight::get\_variance\_residual(model1))

## var.residual   
## 0.5936984

# model 2  
model2\_icc\_3 <- sum(insight::get\_variance\_intercept(model2)) /  
(sum(insight::get\_variance\_intercept(model2)) + insight::get\_variance\_intercept(model2)[2] + insight::get\_variance\_residual(model2))  
  
model2\_icc\_2 <- sum(insight::get\_variance\_intercept(model2)) + insight::get\_variance\_intercept(model2)[2] /  
(sum(insight::get\_variance\_intercept(model2)) + insight::get\_variance\_intercept(model2)[2] + insight::get\_variance\_residual(model2))  
  
  
  
model2\_vpc\_3 <- sum(insight::get\_variance\_intercept(model2)) /  
(sum(insight::get\_variance\_intercept(model2)) + insight::get\_variance\_intercept(model2)[2] + insight::get\_variance\_residual(model2))  
  
model2\_vpc\_2 <- sum(insight::get\_variance\_intercept(model2)[2]) /  
(sum(insight::get\_variance\_intercept(model2)) + insight::get\_variance\_intercept(model2)[2] + insight::get\_variance\_residual(model2))  
  
# model 3  
model3\_icc\_3 <- sum(insight::get\_variance\_intercept(model3)) /  
(sum(insight::get\_variance\_intercept(model3)) + insight::get\_variance\_intercept(model3)[2] + insight::get\_variance\_residual(model3))  
  
model3\_icc\_2 <- sum(insight::get\_variance\_intercept(model3)) + insight::get\_variance\_intercept(model3)[2] /  
(sum(insight::get\_variance\_intercept(model3)) + insight::get\_variance\_intercept(model3)[2] + insight::get\_variance\_residual(model3))  
  
model3\_vpc\_3 <- sum(insight::get\_variance\_intercept(model3)) /  
(sum(insight::get\_variance\_intercept(model3)) + insight::get\_variance\_intercept(model3)[2] + insight::get\_variance\_residual(model3))  
  
model3\_vpc\_2 <- sum(insight::get\_variance\_intercept(model3)[2]) /  
(sum(insight::get\_variance\_intercept(model3)) + insight::get\_variance\_intercept(model3)[2] + insight::get\_variance\_residual(model3))  
  
model2\_icc\_3

## var.intercept.schoolid   
## 0.5756436

model2\_icc\_2

## var.intercept.schoolid   
## 21.36862

model2\_vpc\_3

## var.intercept.schoolid   
## 0.5756436

model2\_vpc\_2

## var.intercept.schoolid   
## 0.02939475

model3\_icc\_3

## var.intercept.schoolid   
## 0.7185651

model3\_icc\_2

## var.intercept.schoolid   
## 21.78368

model3\_vpc\_3

## var.intercept.schoolid   
## 0.7185651

model3\_vpc\_2

## var.intercept.schoolid   
## 0.02167147

performance::icc(model1)

## # Intraclass Correlation Coefficient  
##   
## Adjusted ICC: 0.594  
## Conditional ICC: 0.594

performance::icc(model2)

## # Intraclass Correlation Coefficient  
##   
## Adjusted ICC: 0.593  
## Conditional ICC: 0.593

performance::icc(model3)

## # Intraclass Correlation Coefficient  
##   
## Adjusted ICC: 0.734  
## Conditional ICC: 0.619

table\_5a <- data.frame(Levels = c('School-level', 'Child-level', 'Observation-level'),  
 Model\_1 = c(' ', '21.39', '14.64'),  
 Model\_2 = c('20.25', '1.09', '14.64'),  
 Model\_3 = c('21.11', '.66', '7.87'))  
  
table\_5b\_icc <- data.frame(Levels = c('School-level', 'Child-level'),  
 Model\_1 = c(' ', '.594'),  
 Model\_2 = c('.576', '21.37'),  
 Model\_3 = c('.719', '21.78'))  
  
table\_5c\_vpc <- data.frame(Levels = c('School-level', 'Child-level'),  
 Model\_1 = c(' ', '.594'),  
 Model\_2 = c('.576', '.029'),  
 Model\_3 = c('.719', '.022'))  
  
table\_5a

## Levels Model\_1 Model\_2 Model\_3  
## 1 School-level 20.25 21.11  
## 2 Child-level 21.39 1.09 .66  
## 3 Observation-level 14.64 14.64 7.87

table\_5b\_icc

## Levels Model\_1 Model\_2 Model\_3  
## 1 School-level .576 .719  
## 2 Child-level .594 21.37 21.78

table\_5c\_vpc

## Levels Model\_1 Model\_2 Model\_3  
## 1 School-level .576 .719  
## 2 Child-level .594 .029 .022

1. Answer the following questions using your results from STATA:
2. What does the “beta” for time represent in Models 3, 4 and 5? (What does it explain/capture?)

Answer: the beta for time in all three of these models represents the trajectory of bmi values of the sample as a whole.

1. Briefly describe the VPC results from Models 1, 2 and 3 and what they indicate.

Answer: The VPC results indicate that in the first model, there is a lot of variation found between adolescents. However, when addressing potential differences between adolescents and between schools, more of the variation is shown between schools and even more when including fixed effect variables in the model.

In the first model, 59% of the total variation is attributed to differences found between adolescents. In the second model, 58% of the total variation is attributed to differences between schools, while only 3% of the total variation is attributed to differences between adolescents. In the final model, with the inclusion of fixed effect variables, 72% of the total variation is attributed to differences between schools, while 2% is attributed to differences between adolescents.

1. What does allowing time to be a random variable at levels 2 and 3 (in Model 4) mean?

Answer: By allowing time to be a random variable at levels 2 and 3, we are interested in examining if the association between time and bmi changes by school and by adolescent. We are interested in the change of trajectory of BMI over time and potential differences between schools and between adolescents.

1. In Model 5, explain what allowing “female” to be random at level 3 means. Discuss in general terms and interpret the results from STATA.

Answer: By allowing female to be random at level 3. We are interested in differences in schools between males and females. According to the findings, there was more variation in female BMIs between schools than there was for males.