homework 4

JP

2/3/2020

data <- data %>%   
 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')) %>%   
 mutate(parent\_highestedu = relevel(as.factor(parent\_highestedu), ref = 'college\_degree'),  
 sex = relevel(as.factor(sex), ref = 'male'))

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 thequestions 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 respondentheight and weight in wave 1 of the survey (collected in 1994-95).

§ bmi\_w2 = a continuous measure of body mass index (BMI) calculated from respondentheight and weight in wave 2 of the survey (collected in 1996).

§ bmi\_w3 = a continuous measure of body mass index (BMI) calculated from respondentheight and weight in wave 3 of the survey collected in 2001-2002, when mostrespondents were between ages 18 and 26).

§ bmi\_w4 = a continuous measure of body mass index (BMI) calculated from respondentheight and weight in wave 4 of the survey (collected in 2007-2008, when mostrespondents 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 eitherparent/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

## Tasks:

1. Comparing BMI at two points in time:
2. Obtain summary statistics on BMI in the full sample at each of the four waves(sample size, mean, sd, min, max).

Answer: Below in code.

bmi <- data %>%   
 select(bmi\_w1:bmi\_w4) %>%   
 psych::describe(na.rm = TRUE)  
  
bmi

## vars n mean sd median trimmed mad min max range skew  
## bmi\_w1 1 6291 22.49 4.42 21.50 21.97 3.38 11.22 56.43 45.21 1.50  
## bmi\_w2 2 4698 22.93 4.67 21.82 22.33 3.58 11.95 54.87 42.92 1.57  
## bmi\_w3 3 4757 25.89 5.88 24.43 25.16 4.64 15.19 60.03 44.84 1.41  
## bmi\_w4 4 5026 28.39 6.87 27.08 27.62 5.97 15.98 67.48 51.50 1.22  
## kurtosis se  
## bmi\_w1 3.91 0.06  
## bmi\_w2 3.79 0.07  
## bmi\_w3 2.79 0.09  
## bmi\_w4 2.09 0.10

1. How do these statistics change over time for the sample as a whole?

Answer: The average BMI increases over time.

1. The command “graph matrix variable1 variable2…variable” provides a matrix ofscatter plots. Use this command to obtain a matrix of scatter plots for the fourwaves of BMI data.

plot12 <- data %>%   
 ggplot(aes(bmi\_w1, bmi\_w2)) +  
 geom\_point(color = 'darkgreen') +  
 theme\_minimal()  
  
plot13 <- data %>%   
 ggplot(aes(bmi\_w1, bmi\_w3)) +  
 geom\_point(color = 'darkgreen') +  
 theme\_minimal()  
  
plot14 <- data %>%   
 ggplot(aes(bmi\_w1, bmi\_w4)) +  
 geom\_point(color = 'darkgreen') +  
 theme\_minimal()  
  
plot23 <- data %>%   
 ggplot(aes(bmi\_w2, bmi\_w3)) +  
 geom\_point(color = 'dodgerblue') +  
 theme\_minimal()  
  
plot24 <- data %>%   
 ggplot(aes(bmi\_w2, bmi\_w4)) +  
 geom\_point(color = 'dodgerblue') +  
 theme\_minimal()  
  
plot34 <- data %>%   
 ggplot(aes(bmi\_w3, bmi\_w4)) +  
 geom\_point(color = 'red') +  
 theme\_minimal()  
  
ggpubr::ggarrange(plot12, plot13, plot14,  
 plot23, plot24,  
 plot34,   
 ncol = 2)

## Warning: Removed 1939 rows containing missing values (geom\_point).

## Warning: Removed 1901 rows containing missing values (geom\_point).

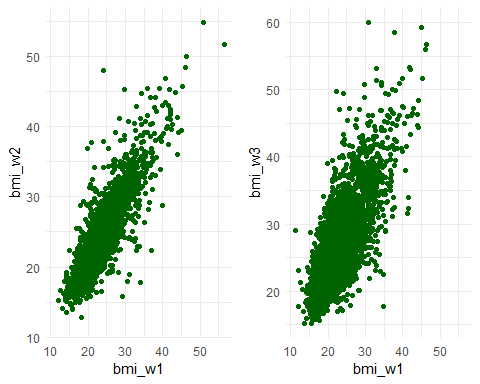
## Warning: Removed 1625 rows containing missing values (geom\_point).

## Warning: Removed 2860 rows containing missing values (geom\_point).

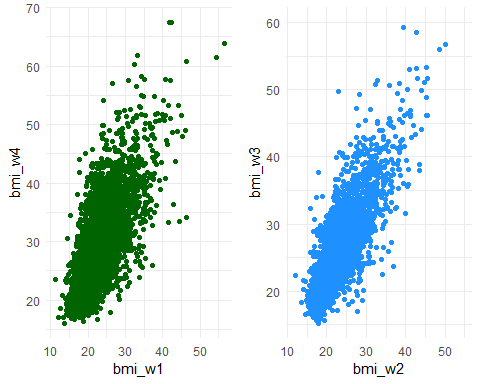
## Warning: Removed 2749 rows containing missing values (geom\_point).

## Warning: Removed 2470 rows containing missing values (geom\_point).

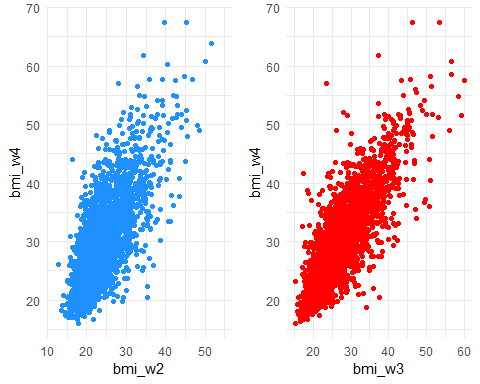
## $`1`



##   
## $`2`



##   
## $`3`



##   
## attr(,"class")  
## [1] "list" "ggarrange"

1. What do you notice, and what does this imply about an individual’s BMI overtime?

Answer: BMI is highly correlated at every time point. It implies that an individual’s BMI will continue to increase over time.

1. In previous assignments we have established that there is school-level clustering of BMI. Using our new knowledge of ICC we will first attempt to quantify this clustering in wave 1.
2. In STATA fit the following Random Intercept models for BMI in wave 1(bmi\_w1):
3. Model 1A: a Null Random Intercepts model with no FE predictors.

ran\_null <- lmer(bmi\_w1 ~ 1 + (1 | schoolid), data = data, REML = FALSE)  
summary(ran\_null)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula: bmi\_w1 ~ 1 + (1 | schoolid)  
## Data: data  
##   
## AIC BIC logLik deviance df.resid   
## 36382.8 36403.1 -18188.4 36376.8 6288   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.6637 -0.6634 -0.2015 0.4359 7.2195   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## schoolid (Intercept) 1.033 1.016   
## Residual 18.511 4.302   
## Number of obs: 6291, groups: schoolid, 132  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 22.346 0.106 125.442 210.8 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

1. Model 2B: a Random Intercept model that adjusts for the following FEpredictors: female gender, age at wave 1, and parent education (collegeor higher is ref category).

ran\_int <- lmer(bmi\_w1 ~ sex + age\_w1 + parent\_highestedu + (1 | schoolid),  
 data = data,  
 REML = FALSE)  
summary(ran\_int)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula: bmi\_w1 ~ sex + age\_w1 + parent\_highestedu + (1 | schoolid)  
## Data: data  
##   
## AIC BIC logLik deviance df.resid   
## 35759.7 35813.6 -17871.9 35743.7 6202   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.5716 -0.6598 -0.2034 0.4250 7.4050   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## schoolid (Intercept) 0.49 0.700   
## Residual 18.20 4.266   
## Number of obs: 6210, groups: schoolid, 132  
##   
## Fixed effects:  
## Estimate Std. Error df t value  
## (Intercept) 16.27495 0.55167 1643.97308 29.501  
## sexfemale -0.26860 0.10905 6174.71821 -2.463  
## age\_w1 0.38547 0.03507 1499.61125 10.991  
## parent\_highesteduhs\_grad 0.41312 0.14641 5378.90950 2.822  
## parent\_highesteduless\_than\_hs 0.94181 0.20577 5190.92772 4.577  
## parent\_highestedusome\_college 0.42366 0.14021 6104.69706 3.022  
## Pr(>|t|)   
## (Intercept) < 0.0000000000000002 \*\*\*  
## sexfemale 0.01380 \*   
## age\_w1 < 0.0000000000000002 \*\*\*  
## parent\_highesteduhs\_grad 0.00479 \*\*   
## parent\_highesteduless\_than\_hs 0.00000483 \*\*\*  
## parent\_highestedusome\_college 0.00252 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) sexfml age\_w1 prnt\_hghstdh\_ prn\_\_\_  
## sexfemale -0.144   
## age\_w1 -0.973 0.044   
## prnt\_hghstdh\_ -0.070 0.002 -0.048   
## prnt\_hghs\_\_ -0.021 0.007 -0.066 0.342   
## prnt\_hghstds\_ -0.096 0.015 -0.023 0.461 0.329

1. For each model in Part A, calculate the ICC (either by hand or using STATA). Ifyou use STATA to generate the ICC, be sure to write the equation STATA usedto calculate it.

sjstats::icc(ran\_null)

## Warning: 'sjstats::icc' is deprecated.  
## Use 'performance::icc()' instead.  
## See help("Deprecated")

## # Intraclass Correlation Coefficient  
##   
## Adjusted ICC: 0.053  
## Conditional ICC: 0.053

sjstats::icc(ran\_int)

## Warning: 'sjstats::icc' is deprecated.  
## Use 'performance::icc()' instead.  
## See help("Deprecated")

## # Intraclass Correlation Coefficient  
##   
## Adjusted ICC: 0.026  
## Conditional ICC: 0.025

# performance::icc() use this in the future

1. How does the ICC change between the two models, and why might this be? Discuss in terms of changing variance parameters at level 1 and level 2.

Answer: The ICC decreased between the two models from .053 to .025. The variance between the models has also decreased between the null model and the adjusted intercept model. The variance and ICC are reduced because of the inclusion of individual-level variables that attributes to the explained variance.

1. What are the two ways we can interpret the calculated ICCs for the random intercepts models?

Answer (Null Model): The expected correlation of outcomes between two individuals in the same group is .053. 5.3% of the total variation is attributed to differences between schools.

Answer (Random Intercept Model): The expected correlation of outcomes between two individuals in the same group while adjusting for fixed values is .025. 2.5% of the total variation is attributed to differences between schools while adjusting for sex, age, and parent’s education.

1. Now fit a Random Slopes Model:
2. Model 3C: Random Slopes model with the same FE predictors as inModel 2, and treat age as a random variable. Allow the covariance to beunstructured. (This isn’t necessarily the “best” sequence to fit models in,but our goal here is just to have a random slopes model to look at).

ran\_slope <- lmer(bmi\_w1 ~ age\_w1 + sex + parent\_highestedu + (age\_w1 | schoolid),  
 data = data,  
 REML = FALSE,  
 control = lmerControl(optimizer ="Nelder\_Mead"))

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl =  
## control$checkConv, : Model failed to converge with max|grad| = 0.00347723  
## (tol = 0.002, component 1)

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable: very large eigenvalue  
## - Rescale variables?

summary(ran\_slope)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula: bmi\_w1 ~ age\_w1 + sex + parent\_highestedu + (age\_w1 | schoolid)  
## Data: data  
## Control: lmerControl(optimizer = "Nelder\_Mead")  
##   
## AIC BIC logLik deviance df.resid   
## 35763.1 35830.4 -17871.6 35743.1 6200   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.5703 -0.6621 -0.2078 0.4216 7.4149   
##   
## Random effects:  
## Groups Name Variance Std.Dev. Corr   
## schoolid (Intercept) 4.22448 2.055   
## age\_w1 0.01464 0.121 -0.94  
## Residual 18.16730 4.262   
## Number of obs: 6210, groups: schoolid, 132  
##   
## Fixed effects:  
## Estimate Std. Error df t value  
## (Intercept) 16.27048 0.58003 101.63232 28.051  
## age\_w1 0.38632 0.03681 93.39144 10.494  
## sexfemale -0.26816 0.10900 6150.52821 -2.460  
## parent\_highesteduhs\_grad 0.41123 0.14636 5354.09186 2.810  
## parent\_highesteduless\_than\_hs 0.94335 0.20571 5107.45861 4.586  
## parent\_highestedusome\_college 0.42442 0.14016 6079.08642 3.028  
## Pr(>|t|)   
## (Intercept) < 0.0000000000000002 \*\*\*  
## age\_w1 < 0.0000000000000002 \*\*\*  
## sexfemale 0.01392 \*   
## parent\_highesteduhs\_grad 0.00498 \*\*   
## parent\_highesteduless\_than\_hs 0.00000463 \*\*\*  
## parent\_highestedusome\_college 0.00247 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) age\_w1 sexfml prnt\_hghstdh\_ prn\_\_\_  
## age\_w1 -0.976   
## sexfemale -0.137 0.042   
## prnt\_hghstdh\_ -0.067 -0.045 0.002   
## prnt\_hghs\_\_ -0.021 -0.062 0.007 0.341   
## prnt\_hghstds\_ -0.093 -0.020 0.015 0.461 0.329  
## convergence code: 0  
## Model failed to converge with max|grad| = 0.00347723 (tol = 0.002, component 1)  
## Model is nearly unidentifiable: very large eigenvalue  
## - Rescale variables?

as.data.frame(VarCorr(ran\_slope))

## grp var1 var2 vcov sdcor  
## 1 schoolid (Intercept) <NA> 4.22447642 2.0553531  
## 2 schoolid age\_w1 <NA> 0.01464395 0.1210122  
## 3 schoolid (Intercept) age\_w1 -0.23455527 -0.9430390  
## 4 Residual <NA> <NA> 18.16730132 4.2623117

# model did not converge due to very large eigenvalue

1. Calculate the ICC for the Random Slope model for individuals with age\_w1 = 18years.

4.22\*1^2 + 2\*-.23\*1\*18 + (.01)\*18^2

## [1] -0.82

slope\_icc <- -.82/(-.82 + 18.17)  
slope\_icc

## [1] -0.04726225

1. How can we interpret the ICC calculated in part F?

Answer: For two individuals at the same school that are both 18 years old, the intraclass correlation coefficient was 4.7%.

1. The primary purpose of this problem is to explore how the importance of schools (i.e., the school-level clustering) for BMI changes over time.
2. Begin by fitting the following Random Intercepts models. As you fit each model ask STATA to give you the ICC. Make note of the ICCs for later.
3. Model 2A: outcome = bmi\_w2, a Null Random Intercepts model with noFE predictors.

ran\_null2 <- lmer(bmi\_w2 ~ 1 + (1 | schoolid), data = data, REML = FALSE)  
summary(ran\_null2)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula: bmi\_w2 ~ 1 + (1 | schoolid)  
## Data: data  
##   
## AIC BIC logLik deviance df.resid   
## 27746.4 27765.7 -13870.2 27740.4 4695   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.6373 -0.6651 -0.2185 0.4089 6.6112   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## schoolid (Intercept) 0.8999 0.9486   
## Residual 20.9382 4.5758   
## Number of obs: 4698, groups: schoolid, 132  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 22.8801 0.1083 126.4812 211.3 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

null2 <- sjstats::icc(ran\_null2)

## Warning: 'sjstats::icc' is deprecated.  
## Use 'performance::icc()' instead.  
## See help("Deprecated")

1. Model 2B: outcome = bmi\_w2, a Random Intercept model that adjusts forthe following FE predictors: female gender, age at wave 1, parenteducation.

ran\_int2 <- lmer(bmi\_w2 ~ sex + age\_w1 + parent\_highestedu + (1 | schoolid),  
 data = data,  
 REML = FALSE)  
summary(ran\_int2)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula: bmi\_w2 ~ sex + age\_w1 + parent\_highestedu + (1 | schoolid)  
## Data: data  
##   
## AIC BIC logLik deviance df.resid   
## 27315.9 27367.5 -13650.0 27299.9 4638   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.6609 -0.6620 -0.2208 0.4002 6.7353   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## schoolid (Intercept) 0.43 0.6557   
## Residual 20.55 4.5332   
## Number of obs: 4646, groups: schoolid, 132  
##   
## Fixed effects:  
## Estimate Std. Error df t value  
## (Intercept) 17.13383 0.71529 1159.22093 23.954  
## sexfemale -0.36005 0.13423 4627.49174 -2.682  
## age\_w1 0.36121 0.04663 1069.25918 7.746  
## parent\_highesteduhs\_grad 0.80828 0.17738 3874.62037 4.557  
## parent\_highesteduless\_than\_hs 1.35702 0.25257 3554.33680 5.373  
## parent\_highestedusome\_college 0.53446 0.17114 4543.76331 3.123  
## Pr(>|t|)   
## (Intercept) < 0.0000000000000002 \*\*\*  
## sexfemale 0.00734 \*\*   
## age\_w1 0.0000000000000219 \*\*\*  
## parent\_highesteduhs\_grad 0.0000053560393461 \*\*\*  
## parent\_highesteduless\_than\_hs 0.0000000825180656 \*\*\*  
## parent\_highestedusome\_college 0.00180 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) sexfml age\_w1 prnt\_hghstdh\_ prn\_\_\_  
## sexfemale -0.155   
## age\_w1 -0.979 0.060   
## prnt\_hghstdh\_ -0.064 -0.014 -0.043   
## prnt\_hghs\_\_ 0.013 -0.002 -0.092 0.327   
## prnt\_hghstds\_ -0.081 0.017 -0.029 0.448 0.316

int2 <- sjstats::icc(ran\_int2)

## Warning: 'sjstats::icc' is deprecated.  
## Use 'performance::icc()' instead.  
## See help("Deprecated")

1. Model 3A: outcome = bmi\_w3, a Null Random Intercepts model with noFE predictors.

ran\_null3 <- lmer(bmi\_w3 ~ 1 + (1 | schoolid), data = data, REML = FALSE)  
summary(ran\_null3)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula: bmi\_w3 ~ 1 + (1 | schoolid)  
## Data: data  
##   
## AIC BIC logLik deviance df.resid   
## 30272.6 30292.0 -15133.3 30266.6 4754   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.0094 -0.6778 -0.2083 0.4538 5.8504   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## schoolid (Intercept) 1.426 1.194   
## Residual 33.103 5.754   
## Number of obs: 4757, groups: schoolid, 132  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 25.8436 0.1364 133.4439 189.5 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

null3 <- sjstats::icc(ran\_null3)

## Warning: 'sjstats::icc' is deprecated.  
## Use 'performance::icc()' instead.  
## See help("Deprecated")

1. Model 3B: outcome = bmi\_w3, a Random Intercept model that adjusts forthe following FE predictors: female gender, age at wave 1, parenteducation.

ran\_int3 <- lmer(bmi\_w3 ~ sex + age\_w1 + parent\_highestedu + (1 | schoolid),  
 data = data,  
 REML = FALSE)  
summary(ran\_int3)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula: bmi\_w3 ~ sex + age\_w1 + parent\_highestedu + (1 | schoolid)  
## Data: data  
##   
## AIC BIC logLik deviance df.resid   
## 29865.5 29917.2 -14924.8 29849.5 4697   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.1198 -0.6770 -0.2147 0.4504 5.8405   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## schoolid (Intercept) 0.9647 0.9822   
## Residual 32.6834 5.7169   
## Number of obs: 4705, groups: schoolid, 132  
##   
## Fixed effects:  
## Estimate Std. Error df t value  
## (Intercept) 21.61036 0.83552 1394.15814 25.865  
## sexfemale -0.17529 0.16844 4669.97198 -1.041  
## age\_w1 0.23864 0.05333 1262.67535 4.475  
## parent\_highesteduhs\_grad 1.13137 0.22453 4063.55515 5.039  
## parent\_highesteduless\_than\_hs 1.48552 0.32440 3957.18513 4.579  
## parent\_highestedusome\_college 0.82556 0.21510 4630.72492 3.838  
## Pr(>|t|)   
## (Intercept) < 0.0000000000000002 \*\*\*  
## sexfemale 0.298083   
## age\_w1 0.000008348 \*\*\*  
## parent\_highesteduhs\_grad 0.000000488 \*\*\*  
## parent\_highesteduless\_than\_hs 0.000004811 \*\*\*  
## parent\_highestedusome\_college 0.000126 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) sexfml age\_w1 prnt\_hghstdh\_ prn\_\_\_  
## sexfemale -0.160   
## age\_w1 -0.973 0.054   
## prnt\_hghstdh\_ -0.058 -0.011 -0.057   
## prnt\_hghs\_\_ -0.031 0.006 -0.053 0.323   
## prnt\_hghstds\_ -0.088 0.008 -0.029 0.446 0.309

int3 <- sjstats::icc(ran\_int3)

## Warning: 'sjstats::icc' is deprecated.  
## Use 'performance::icc()' instead.  
## See help("Deprecated")

1. Model 4A: outcome = bmi\_w4, a Null Random Intercepts model with no FE predictors.

ran\_null4 <- lmer(bmi\_w4 ~ 1 + (1 | schoolid), data = data, REML = FALSE)  
summary(ran\_null4)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula: bmi\_w4 ~ 1 + (1 | schoolid)  
## Data: data  
##   
## AIC BIC logLik deviance df.resid   
## 33567.3 33586.9 -16780.6 33561.3 5023   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.9092 -0.7032 -0.1898 0.5114 5.6824   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## schoolid (Intercept) 1.55 1.245   
## Residual 45.54 6.748   
## Number of obs: 5026, groups: schoolid, 132  
##   
## Fixed effects:  
## Estimate Std. Error df t value Pr(>|t|)   
## (Intercept) 28.3422 0.1479 134.4399 191.6 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

null4 <- sjstats::icc(ran\_null4)

## Warning: 'sjstats::icc' is deprecated.  
## Use 'performance::icc()' instead.  
## See help("Deprecated")

1. Model 4B: outcome = bmi\_w4, a Random Intercept model that adjusts forthe following FE predictors: female gender, age at wave 1, parenteducation.

ran\_int4 <- lmer(bmi\_w4 ~ sex + age\_w1 + parent\_highestedu + (1 | schoolid),  
 data = data,  
 REML = FALSE)  
summary(ran\_int4)

## Linear mixed model fit by maximum likelihood . t-tests use  
## Satterthwaite's method [lmerModLmerTest]  
## Formula: bmi\_w4 ~ sex + age\_w1 + parent\_highestedu + (1 | schoolid)  
## Data: data  
##   
## AIC BIC logLik deviance df.resid   
## 33057.7 33109.8 -16520.8 33041.7 4951   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -2.0626 -0.7012 -0.1872 0.5100 5.6874   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## schoolid (Intercept) 1.077 1.038   
## Residual 45.088 6.715   
## Number of obs: 4959, groups: schoolid, 132  
##   
## Fixed effects:  
## Estimate Std. Error df t value  
## (Intercept) 24.63889 0.95783 1429.08805 25.724  
## sexfemale 0.23535 0.19281 4939.96366 1.221  
## age\_w1 0.17619 0.06093 1274.22633 2.892  
## parent\_highesteduhs\_grad 1.49835 0.25600 4242.99475 5.853  
## parent\_highesteduless\_than\_hs 1.84414 0.36430 4125.00095 5.062  
## parent\_highestedusome\_college 1.08060 0.24490 4869.39220 4.413  
## Pr(>|t|)   
## (Intercept) < 0.0000000000000002 \*\*\*  
## sexfemale 0.22229   
## age\_w1 0.00389 \*\*   
## parent\_highesteduhs\_grad 0.0000000052 \*\*\*  
## parent\_highesteduless\_than\_hs 0.0000004327 \*\*\*  
## parent\_highestedusome\_college 0.0000104380 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) sexfml age\_w1 prnt\_hghstdh\_ prn\_\_\_  
## sexfemale -0.160   
## age\_w1 -0.974 0.054   
## prnt\_hghstdh\_ -0.067 -0.010 -0.048   
## prnt\_hghs\_\_ -0.028 -0.005 -0.056 0.327   
## prnt\_hghstds\_ -0.097 0.011 -0.021 0.451 0.316

int4 <- sjstats::icc(ran\_int4)

## Warning: 'sjstats::icc' is deprecated.  
## Use 'performance::icc()' instead.  
## See help("Deprecated")

1. Make a table to organize your results for how ICC changes over time. Use something like the following setup:

null2

## # Intraclass Correlation Coefficient  
##   
## Adjusted ICC: 0.041  
## Conditional ICC: 0.041

int2

## # Intraclass Correlation Coefficient  
##   
## Adjusted ICC: 0.020  
## Conditional ICC: 0.020

null3

## # Intraclass Correlation Coefficient  
##   
## Adjusted ICC: 0.041  
## Conditional ICC: 0.041

int3

## # Intraclass Correlation Coefficient  
##   
## Adjusted ICC: 0.029  
## Conditional ICC: 0.028

null4

## # Intraclass Correlation Coefficient  
##   
## Adjusted ICC: 0.033  
## Conditional ICC: 0.033

int4

## # Intraclass Correlation Coefficient  
##   
## Adjusted ICC: 0.023  
## Conditional ICC: 0.023

table <- data.frame(Variable = c('Null Model ICC',  
 'Adjusted Model ICC'),  
 wave\_1\_bmi = c('.053', '.025'),  
 wave\_2\_bmi = c('.041', '.020'),  
 wave\_3\_bmi = c('.041', '.028'),  
 wave\_4\_bmi = c('.033', '.023'))  
  
table

## Variable wave\_1\_bmi wave\_2\_bmi wave\_3\_bmi wave\_4\_bmi  
## 1 Null Model ICC .053 .041 .041 .033  
## 2 Adjusted Model ICC .025 .020 .028 .023

1. How does the ICC change over time? What does this say about persistence of school effects?

Answer: The ICC for the null models changes slightly from 5% to 3% by wave 4. However, the adjusted model reveals that the inclusion of sex, age, and parent’s highest education makes the ICC remain constant across the ways. The persistence of school effects is that that there is some effect that is due to school differences when including fixed effects on bmi at different time points.

1. This problem is designed to help you get started on the code you will use in your finalproject. If you are electing NOT to complete the final project and will instead take theexam, use this problem as a chance to think about a hypothetical project you might liketo do. Note that in this problem you will be developing STATA code, but are not going tobe required to run any of it. (I realize many of you won’t have the data you are usingcleaned and ready to use yet.)
2. What is the basic data structure you will be working with? (e.g., two-level model,three-level…?). Using Goldstein notation, draw this data structure.

Answer:

State

County

1. What is your primary outcome of interest? (While you may have interest inseveral outcomes, for the final project you will only be required to use one ofthem. If possible I recommend using a continuous outcome because that is thefocus of our class. You may also use other types of outcomes, and we will coverlogistic/Poisson models in class at a later date).

Answer: The outcome of interest is a measure of physical inactivity rates. It is continuous.

1. What are your primary predictors of interest (FE or RE variables), and what levelis each one associated with?

Answer: Fixed effects/adjusted variables will be population, adult obesity, percent rurality, violent crime, median household income. All of these variables are county-level (individual-level)

Random effects will be access to exercise opportunities over the state-level (between-group).

1. Write the code you will use to determine:
2. How many observations are in your sample?

data %>%   
 count()

1. How many groups (level 2 units) are in your sample?

data %>%   
 group\_by(state\_fips\_code) %>%   
 summarize(n = n()) %>%   
 ungroup() %>%   
 count(state\_fips\_code) %>%   
 summarize(states = sum(n))

1. The number of level 1 units in level 2 groups – don’t print out the samplesize for each group, but rather describe the distribution of group sizes inthe sample (ex: what is the average number of adolescents per school?What is the min/max number of adolescents per school?).

data %>%   
 group\_by(state\_fips\_code) %>%   
 mutate(county\_per = n()) %>%   
sundry::descrips(county\_per)

* Hint: go back to the slides and look for the “egen” command and the“generate tolist” code. These may be helpful.

1. Now make a list of the models you will want to fit. These can include single-levelmodels, multilevel random intercept models, and multilevel random slope models.

null\_state <- lmer(physical\_inactivity ~ 1 +  
 (1 | state\_fips\_code), data = full\_reduced\_zz)  
summary(null\_state)  
sjstats::icc(null\_state)  
  
  
random\_int\_state <- lmer(physical\_inactivity ~ population + adult\_obesity +  
 percent\_rural +   
 violent\_crime + median\_household\_income + 1 +  
 (1 | state\_fips\_code), data = data)  
summary(random\_int\_state)  
sjstats::icc(random\_int\_state)  
  
  
state\_no\_inter\_no\_vcov <- lmer(physical\_inactivity ~ population + adult\_obesity +  
 percent\_rural +   
 violent\_crime + median\_household\_income +   
 access\_to\_exercise\_opportunities +  
 (access\_to\_exercise\_opportunities || state\_fips\_code),   
 data = data,  
 control = lmerControl(optimizer ="Nelder\_Mead"))  
summary(state\_no\_inter\_no\_vcov)  
as.data.frame(VarCorr(state\_no\_inter\_no\_vcov))  
# calculate icc for random slope model

Fixed effect: adult obesity, population, rurality, violent crime, median household income

Random effect: access to exercise opportunities

For each model, be sure to identify how different variables will be used (e.g., “asa FE predictor” or “as a random slope variable”). You do not need to write theactual models (equations) for this, because some of you may be interested inmodels we haven’t discussed yet in class. The goal here is to develop the “gameplan” for your project.

Extra Credit: (worth 1 point)

§ Review the article “Evans et al-jech-2015” posted under “Articles of Interest” on the Filespage of Canvas. This was a short report we wrote as a “hypothesis generating” piece,and it was conducted in the full Add Health sample. The analyses in this paperresemble the analyses you conducted above.

§ Answer the following questions:

o What were the general findings from the gender-stratified and race/ethnicitystratified versions of the models?

Answer: For sex-stratified findings, the results were similar, females’ and males’ BMI increased over time when adjusting for age, race, and parental education. For race/ethnicity-stratified findings, there was clustering by school among all racial/ethnic groups (Black, Hispanic, White, Other), but the total variation of BMI attributable to schools was greatest in Whites.

o What are some of the hypothesized explanations for why clustering at the wave 1school-level persists into adulthood? In other words, why might adult individualscontinue to have BMIs correlated with the BMIs of those they went to school withas adolescents?

Answer: One potential reason is that schools are a salient factor in shaping the health behaviors that affect health behaviors in adulthood, which may then result in increased obesity. The other explanation is that young adults select friends that may resemble individuals that they were around in school. Then these behaviors still exist in these new adult peer groups.