2480 Final Project

Yi Yang

2024-04-11

Package Upload

```r
library(readr)
library(haven)
library(psych)
library(tidyverse)
library(labelled)
library(table1)
library(dplyr)
library(haven)
library(tidyverse)
library(ggplot2)
library(lme4)
library(broom)
library(naniar)
library(sjPlot)
library(labelled)
library(performance)
library(knitr)
library(kableExtra)
library(lmerTest)
library(pander)
library(performance)
library(corrplot)
```

Upload Data

```r
data <- read_dta("finalproj_2023.dta")
head(data)
```

```
## # A tibble: 6 x 320
##     PID   TAS TAS05 TAS07 TAS09 TAS11 TAS13 TAS15 TAS17 TAS19 ER30000    ER30001
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl+lbl>    <dbl>
## 1  4037     1    NA    NA     1    NA    NA    NA    NA    NA 3 [Releas~       4
## 2  4038     2    NA    NA     1     1    NA    NA    NA    NA 3 [Releas~       4
## 3  4039     5    NA    NA     1     1     1     1     1    NA 3 [Releas~       4
## 4  4041     5    NA    NA    NA     1     1     1     1     1 3 [Releas~       4
## 5  4042     1    NA    NA    NA    NA    NA    NA    NA     1 3 [Releas~       4
## 6  4180     4     1     1     1     1    NA    NA    NA    NA 3 [Releas~       4
```

```
## # ... with 308 more variables: ER30002 <dbl>, ER33801 <dbl>, ER33802 <dbl+lbl>,
## #   ER33803 <dbl+lbl>, ER33804 <dbl>, TA050001 <dbl+lbl>, TA050078 <dbl+lbl>,
## #   TA050676 <dbl+lbl>, TA050679 <dbl+lbl>, TA050686 <dbl+lbl>,
## #   TA050690 <dbl+lbl>, TA050693 <dbl+lbl>, TA050708 <dbl+lbl>,
## #   TA050720 <dbl+lbl>, TA050762 <dbl+lbl>, TA050766 <dbl+lbl>,
## #   TA050770 <dbl+lbl>, TA050778 <dbl+lbl>, TA050786 <dbl+lbl>,
## #   TA050790 <dbl+lbl>, TA050794 <dbl+lbl>, TA050802 <dbl+lbl>, ...
```

Data Cleaning

```r
data <- data %>%
 mutate(PID = (ER30001 * 1000) + ER30002) %>%
 relocate(PID) #putting at beginning of dataset
obs <- dim(data)[1]
obs
```

```
## [1] 4776
```

```r
sum(duplicated(data$PID))
```

```
## [1] 0
```

```r
data$PID <- as.integer(data$PID)
data$Anxiety1<- data$TA050933
data$Anxiety2<- data$TA070914
data$Anxiety3<- data$TA090978
data$Anxiety4<- data$TA111120
data$Anxiety5<- data$TA131212
data$Smoking <- data$TA050762
data$Race <- data$TA050884
data$age1 <- data$ER33804
data$age2 <- data$ER33904
data$age3 <- data$ER34004
data$age4 <- data$ER34104
data$age5 <- data$ER34204


sample_dat <- data %>%
  select(Anxiety1,Anxiety2,Anxiety3,Anxiety4,Anxiety5,
         Smoking, Race, age1,age2,age3,age4,age5,PID) %>%
  dplyr::mutate(Race = case_when(
    Race == 1 ~ "White",
    Race == 2 ~ "Black",
    Race == 3 ~ "Other",
    Race == 4 ~ "Other",
    Race == 5 ~ "Other",
    Race == 7 | Race == 8 | Race == 9 ~ NA_character_
  ))
table(data$Race, useNA = "always")
```

```
##
##    1    2    3    4    5    7    8    9 <NA>
##  378  312    6    8    3    8    2   28 4031
```

```
sample_dat_1 <- sample_dat %>% filter(!is.na(Anxiety1))
dim(sample_dat_1)
```

```
## [1] 745  13
```

```
sample_dat_2 <- sample_dat_1 %>% filter(!is.na(Anxiety2))
sample_dat_3 <- sample_dat_2 %>% filter(!is.na(Anxiety3))
sample_dat_4 <- sample_dat_3 %>% filter(!is.na(Anxiety4))
sample_dat_5 <- sample_dat_4 %>% filter(!is.na(Anxiety5))
sample_dat_6 <- sample_dat_5 %>% filter(!is.na(Smoking))
dim(sample_dat_6)
```

```
## [1] 238  13
```

```
table1(~.|Anxiety1 , data = sample_dat_6)
```

```
## Warning in table1.formula(~. | Anxiety1, data = sample_dat_6): Terms to the
## right of '|' in formula 'x' define table columns and are expected to be factors
## with meaningful labels.
```

	1	2	3	4
	(N=14)	(N=44)	(N=61)	(N=54)
MENTAL HEALTH: SOCIAL ANXIETY				
Mean (SD)	1.79 (1.12)	2.73 (1.30)	2.97 (0.948)	3.76 (1.23)
Median [Min, Max]	1.50 [1.00, 5.00]	2.50 [1.00, 7.00]	3.00 [1.00, 5.00]	4.00 [2.00, 6.
MENTAL HEALTH: SOCIAL ANXIETY				
Mean (SD)	1.64 (0.633)	2.75 (1.33)	2.97 (1.14)	3.83 (1.41)
Median [Min, Max]	2.00 [1.00, 3.00]	3.00 [1.00, 7.00]	3.00 [1.00, 6.00]	4.00 [1.00, 7.
MENTAL HEALTH: SOCIAL ANXIETY				
Mean (SD)	2.21 (1.42)	2.77 (1.44)	2.90 (1.14)	3.44 (1.21)
Median [Min, Max]	2.00 [1.00, 6.00]	2.50 [1.00, 6.00]	3.00 [1.00, 6.00]	3.00 [1.00, 6.
MENTAL HEALTH: SOCIAL ANXIETY				
Mean (SD)	2.07 (0.917)	2.93 (1.45)	3.02 (1.18)	3.61 (1.28)
Median [Min, Max]	2.00 [1.00, 4.00]	3.00 [1.00, 6.00]	3.00 [1.00, 6.00]	4.00 [1.00, 6.
H32 WTR EVER SMOKED CIGARETTES				
Mean (SD)	3.29 (2.40)	3.07 (2.37)	3.48 (2.22)	4.13 (1.75)
Median [Min, Max]	5.00 [0, 5.00]	5.00 [0, 5.00]	5.00 [0, 5.00]	5.00 [0, 5.00]
Race				
Black	8 (57.1%)	17 (38.6%)	19 (31.1%)	24 (44.4%)
White	6 (42.9%)	21 (47.7%)	35 (57.4%)	28 (51.9%)
Other	0 (0%)	2 (4.5%)	1 (1.6%)	1 (1.9%)
Missing	0 (0%)	4 (9.1%)	6 (9.8%)	1 (1.9%)
AGE OF INDIVIDUAL 05				
Mean (SD)	17.9 (0.730)	17.9 (0.563)	18.0 (0.617)	18.1 (0.627)
Median [Min, Max]	18.0 [17.0, 19.0]	18.0 [17.0, 19.0]	18.0 [17.0, 20.0]	18.0 [17.0, 1
AGE OF INDIVIDUAL 07				
Mean (SD)	20.1 (0.829)	20.0 (0.480)	20.0 (0.617)	20.1 (0.640)
Median [Min, Max]	20.0 [19.0, 21.0]	20.0 [19.0, 21.0]	20.0 [19.0, 22.0]	20.0 [19.0, 2
AGE OF INDIVIDUAL 09				
Mean (SD)	21.9 (0.730)	21.9 (0.520)	21.9 (0.569)	22.0 (0.629)
Median [Min, Max]	22.0 [21.0, 23.0]	22.0 [21.0, 23.0]	22.0 [21.0, 23.0]	22.0 [21.0, 2
AGE OF INDIVIDUAL 11				
Mean (SD)	24.0 (0.784)	24.0 (0.608)	24.0 (0.576)	24.0 (0.643)
Median [Min, Max]	24.0 [23.0, 25.0]	24.0 [23.0, 25.0]	24.0 [23.0, 25.0]	24.0 [23.0, 2
AGE OF INDIVIDUAL 13				
Mean (SD)	26.0 (0.784)	25.9 (0.493)	25.9 (0.629)	26.0 (0.582)
Median [Min, Max]	26.0 [25.0, 27.0]	26.0 [25.0, 27.0]	26.0 [25.0, 27.0]	26.0 [25.0, 2
PID				
Mean (SD)	3900000 (1950000)	3180000 (2210000)	3120000 (1990000)	3100000 (228
Median [Min, Max]	4060000 [1620000, 6550000]	2750000 [53000, 6830000]	2300000 [89000, 6660000]	2230000 [173

3.a Descriptive Statistics of the data

```r
head(sample_dat_6)
```

```
## # A tibble: 6 x 13
##   Anxiety1 Anxie~1 Anxie~2 Anxie~3 Anxie~4 Smoking Race   age1  age2  age3  age4
```

```
##   <dbl+lb> <dbl+l> <dbl+l> <dbl+l> <dbl+l> <dbl+l> <chr> <dbl> <dbl> <dbl> <dbl>
## 1 2 [Actu~ 4 [Act~ 3 [Act~ 4 [Act~ 6 [Act~ 0 [Ina~ White    18    20    22    24
## 2 2 [Actu~ 2 [Act~ 1 [Act~ 1 [Act~ 1 [Act~ 5 [No]  White    19    20    23    25
## 3 2 [Actu~ 2 [Act~ 3 [Act~ 3 [Act~ 4 [Act~ 1 [Yes] White    18    20    22    24
## 4 3 [Actu~ 4 [Act~ 3 [Act~ 3 [Act~ 4 [Act~ 5 [No]  White    17    19    21    23
## 5 6 [Actu~ 6 [Act~ 5 [Act~ 2 [Act~ 2 [Act~ 1 [Yes] White    19    21    23    25
## 6 2 [Actu~ 2 [Act~ 3 [Act~ 3 [Act~ 3 [Act~ 0 [Ina~ White    18    20    22    24
## # ... with 2 more variables: age5 <dbl>, PID <int>, and abbreviated variable
## #   names 1: Anxiety2, 2: Anxiety3, 3: Anxiety4, 4: Anxiety5
```

```
dim(sample_dat_6)
```

```
## [1] 238  13
```

```
des<-sample_dat_6 %>% describe()
# Descriptive statistics of the data
des
```

```
##           vars   n       mean         sd    median    trimmed        mad   min
## Anxiety1     1 238       3.61       1.45       3.5       3.57       2.22     1
## Anxiety2     2 238       3.49       1.48       3.0       3.43       1.48     1
## Anxiety3     3 238       3.41       1.51       3.0       3.33       1.48     1
## Anxiety4     4 238       3.32       1.43       3.0       3.29       1.48     1
## Anxiety5     5 238       3.34       1.40       3.0       3.32       1.48     1
## Smoking      6 238       3.53       2.17       5.0       3.78       0.00     0
## Race*        7 224       2.13       0.98       3.0       2.17       0.00     1
## age1         8 238      17.99       0.62      18.0      17.98       0.00    17
## age2         9 238      20.04       0.62      20.0      20.04       0.00    19
## age3        10 238      21.98       0.60      22.0      21.98       0.00    21
## age4        11 238      23.99       0.63      24.0      23.99       0.00    23
## age5        12 238      26.00       0.61      26.0      25.99       0.00    25
## PID         13 238 3231879.35 2118429.30 2628535.5 3177827.92 2513011.45 53036
##               max   range  skew kurtosis        se
## Anxiety1        7       6  0.21    -0.69      0.09
## Anxiety2        7       6  0.30    -0.67      0.10
## Anxiety3        7       6  0.47    -0.43      0.10
## Anxiety4        7       6  0.26    -0.72      0.09
## Anxiety5        7       6  0.24    -0.42      0.09
## Smoking         5       5 -0.82    -1.27      0.14
## Race*           3       2 -0.27    -1.92      0.07
## age1           20       3  0.11    -0.02      0.04
## age2           22       3  0.08    -0.09      0.04
## age3           23       2  0.01    -0.26      0.04
## age4           25       2  0.01    -0.49      0.04
## age5           27       2  0.00    -0.29      0.04
## PID       6857184 6804148  0.30    -1.37 137317.38
```

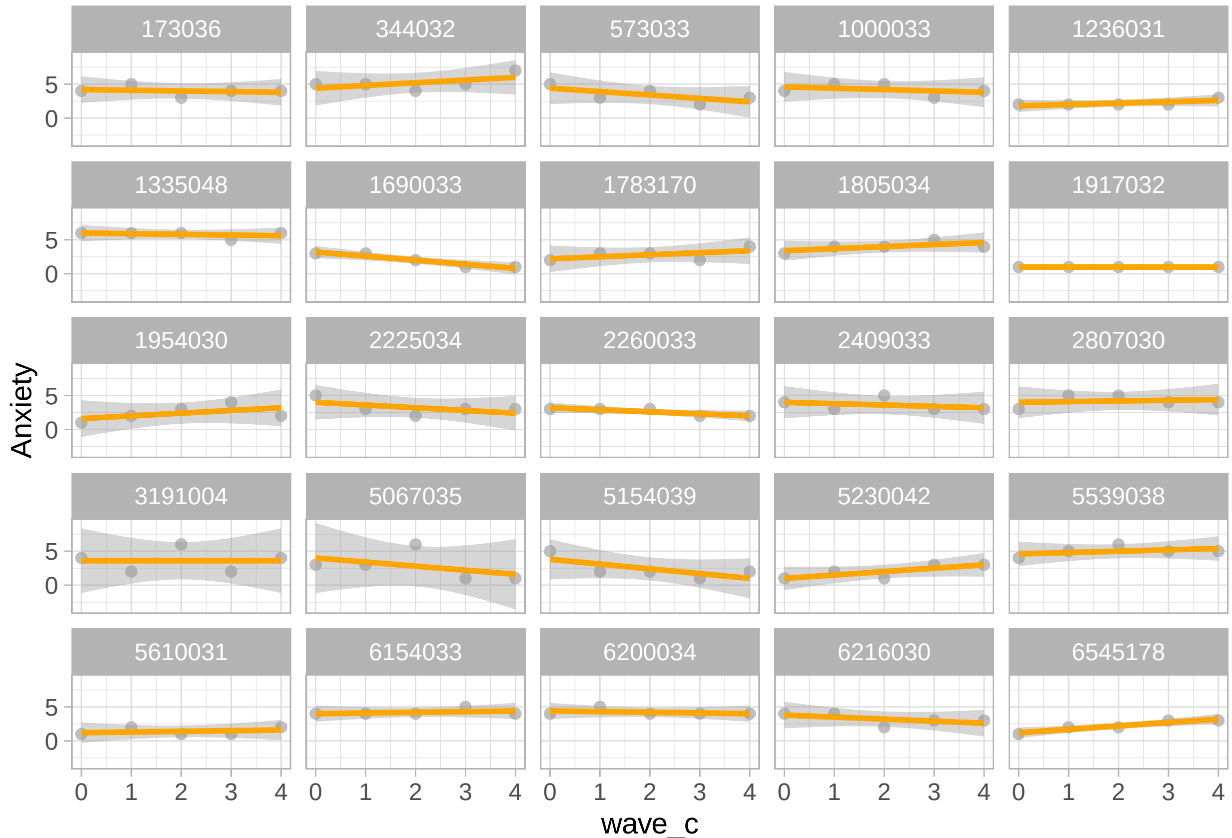
```
# Calculate the correlation coefficient matrix for Anxiety
cor_anx<-cor(sample_dat_6[c("Anxiety1","Anxiety2","Anxiety3","Anxiety4", "Anxiety5")],
             use = "pairwise.complete.obs" )
# correlation coefficient matrix for Anxiety
cor_anx
```

```
##           Anxiety1  Anxiety2  Anxiety3  Anxiety4  Anxiety5
## Anxiety1 1.0000000 0.6026319 0.5040263 0.4473227 0.3735909
## Anxiety2 0.6026319 1.0000000 0.5455061 0.4900145 0.4933885
## Anxiety3 0.5040263 0.5455061 1.0000000 0.5960024 0.4770776
## Anxiety4 0.4473227 0.4900145 0.5960024 1.0000000 0.6190086
## Anxiety5 0.3735909 0.4933885 0.4770776 0.6190086 1.0000000
```

3.b i Describe the growth in your outcome

```r
obs <- dim(sample_dat_6)[1] # data size
set.seed(0)
sample_data <- sample_dat_6[sample(obs, size = 25),] # Sampling of 25 samples
sample_dat_long <- sample_data %>%
  select(Anxiety1,Anxiety2,Anxiety3,Anxiety4,Anxiety5,PID,Race) %>%
  pivot_longer(cols = c("Anxiety1","Anxiety2","Anxiety3","Anxiety4","Anxiety5"),
               values_to = "Anxiety") %>% mutate(wave = case_when(
                 name == "Anxiety1" ~ 1,
                 name == "Anxiety2" ~ 2,
                 name == "Anxiety3" ~ 3,
                 name == "Anxiety4" ~ 4,
                 name == "Anxiety5" ~ 5))
sample_dat_long$wave_c <- (sample_dat_long$wave) - 1
#Individual growth plots
ggplot(data = sample_dat_long, aes(x = wave_c, y = Anxiety)) +
  geom_point(col='gray') + geom_smooth(method = "lm",col='orange') +
  facet_wrap(vars(PID))+theme_light()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

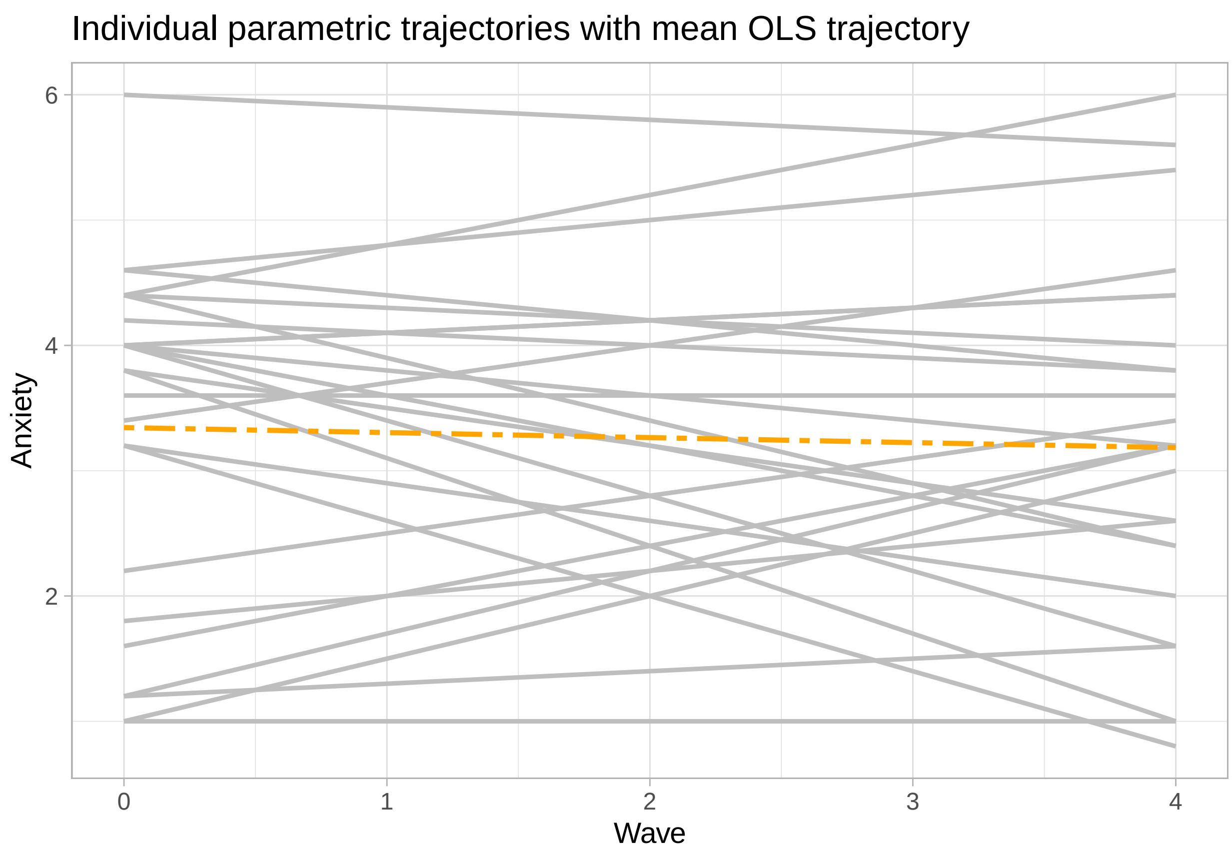


3.b ii Individual OLS regressions conducted and visualized with the mean trajectory line.

```
#Individual parametric trajectories with mean OLS trajectory
ggplot(data = sample_dat_long, aes(x = wave_c, y = Anxiety)) +
  geom_smooth(aes(group = as.factor(PID)), method = "lm", color="gray",cex=0.8,se=F) +
  geom_smooth(method = "lm",color ="orange",se=F,cex=0.9,lty=6)+
  labs(x="Wave",y="Anxiety",title="Individual parametric trajectories with mean OLS trajectory")+
  theme_light()
```

```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```

```
## ‘geom_smooth()‘ using formula = ’y ~ x’
## ‘geom_smooth()‘ using formula = ’y ~ x’
```



3.c i sample means of the estimated intercepts and slopes

```
sample_dat_long <- sample_dat %>%
  select(Anxiety1,Anxiety2,Anxiety3,Anxiety4,Anxiety5,PID) %>%
  pivot_longer(cols = c("Anxiety1","Anxiety2","Anxiety3","Anxiety4","Anxiety5"),
               values_to = "Anxiety") %>% mutate(wave = case_when(
                 name == "Anxiety1" ~ 1,
                 name == "Anxiety2" ~ 2,
                 name == "Anxiety3" ~ 3,
                 name == "Anxiety4" ~ 4,
                 name == "Anxiety5" ~ 5))
sample_dat_long$wave_c <- sample_dat_long$wave - 1

# Group by PID and create a new missing wave column
sample_dat_long_2 <- sample_dat_long %>%
  group_by(PID) %>%
  dplyr::mutate(missing_waves = sum(is.na(Anxiety)))

# Group by PID and filter for missing_wave less than 3
sample_dat_long3 <- sample_dat_long_2 %>%
  group_by(PID) %>%
  filter(sum(missing_waves) < 3)

# Building a linear model
```

```r
model1 <- sample_dat_long3  %>% dplyr::group_by(PID) %>%
  do(model = lm(Anxiety ~ wave_c, data =.))
model1[[2]][[1]]
```

```
##
## Call:
## lm(formula = Anxiety ~ wave_c, data = .)
##
## Coefficients:
## (Intercept)       wave_c
##         2.2          0.8
```

```r
intercept <- slope <- NULL

# Calling slope and intercept
for(i in 1:nrow(model1)){
  intercept[i] <- model1[[2]][[i]][["coefficients"]][1]
  slope[i] <- model1[[2]][[i]][["coefficients"]][2]
}
```

3.c ii Sample Variance

```r
summary(intercept)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.600   2.600   3.400   3.575   4.400   6.800
```

```r
var(intercept)
```

```
## [1] 1.932864
```

```r
summary(slope)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -1.20000 -0.30000 -0.10000 -0.07017  0.17500  1.10000
```

```r
var(slope)
```

```
## [1] 0.1469122
```

3.c iii correlation between the estimated intercepts and slopes

```r
# Check the covariance of slope and intercept
cor(intercept,slope)
```

```
## [1] -0.5745468
```

3.D Model building

3.D.i Conduct the unconditional mean model

```r
dat_long <- sample_dat %>%
  dplyr::select(Anxiety1,Anxiety2,Anxiety3,Anxiety4,Anxiety5,
                age1,age2,age3,age4,age5,PID)  %>%
  pivot_longer(-PID) %>%
  separate(name, into = c("name", "wave"), sep = "(?<=[A-Za-z])(?=[0-9])") %>%
  pivot_wider(names_from = "name", values_from = "value")

dat_mari_race <- sample_dat %>%
  select(Smoking, Race,PID)

dat_long <- left_join(dat_long, dat_mari_race, by = "PID")

dat_long <- remove_labels(dat_long)

dat_long$wave_c <- as.integer(dat_long$wave)-1
table(dat_long$Race)
```

3.D.i 1 Interpret the fixed and random effects

```
##
## Black Other White
##  1560    85  1890
```

```r
model.a <- lmer(Anxiety ~ 1 + (1 |PID), data = dat_long, REML = FALSE)
summary(model.a)
```

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: Anxiety ~ 1 + (1 | PID)
##    Data: dat_long
##
##      AIC      BIC   logLik deviance df.resid
##  23929.8  23950.4 -11961.9  23923.8     7122
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1412 -0.5761 -0.0707  0.5550  3.9357
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  PID      (Intercept) 1.274    1.129
##  Residual             1.003    1.001
## Number of obs: 7125, groups:  PID, 2570
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 3.431e+00  2.587e-02 2.508e+03   132.6   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
performance::icc(model.a)
```

3.D.i 2Conduct the ICC and interpret

```
## # Intraclass Correlation Coefficient
##
##     Adjusted ICC: 0.559
##   Unadjusted ICC: 0.559
```

```r
icc_n <- as.data.frame(VarCorr(model.a),comp="Variance")$vcov[1]
icc_d <- as.data.frame(VarCorr(model.a),comp="Variance")$vcov[1] +
         as.data.frame(VarCorr(model.a),comp="Variance")$vcov[2]
icc_n / icc_d
```

```
## [1] 0.5594694
```

3.D.ii Conduct the unconditional growth model

```r
model.b <- lmer(Anxiety ~ wave_c + (wave_c|PID), data = dat_long, REML = FALSE)
summary(model.b)
```

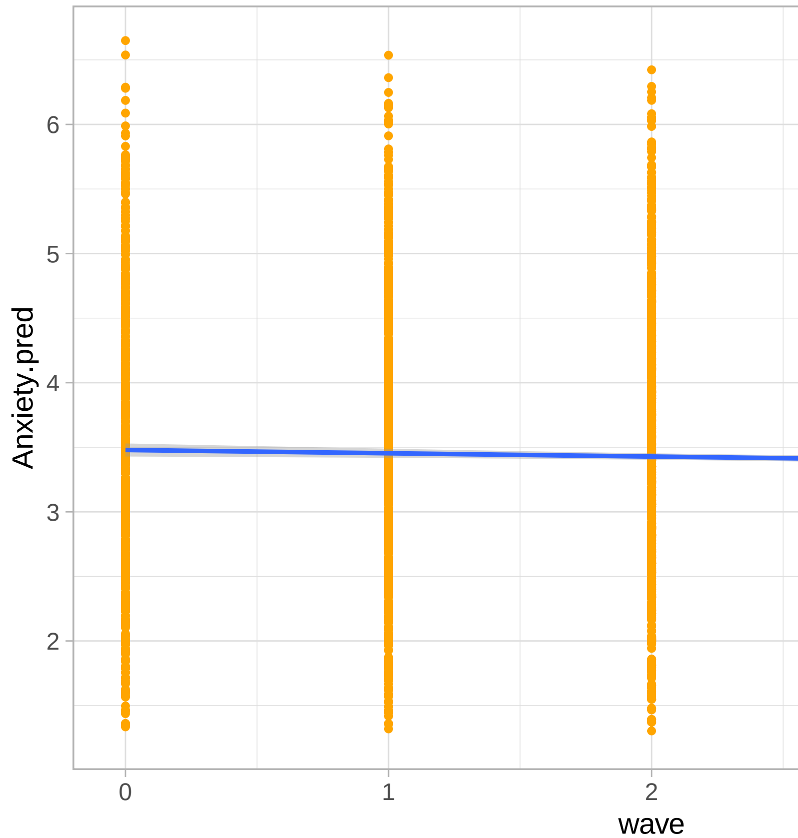
3.D.ii 1 Interpret the fixed and random effects

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: Anxiety ~ wave_c + (wave_c | PID)
##    Data: dat_long
##
##      AIC      BIC   logLik deviance df.resid
##  23865.5  23906.8 -11926.8  23853.5     7119
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3626 -0.5509 -0.0683  0.5270  4.1238
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  PID      (Intercept) 1.56071  1.2493
##           wave_c      0.05378  0.2319   -0.42
##  Residual             0.90482  0.9512
## Number of obs: 7125, groups:  PID, 2570
##
## Fixed effects:
##              Estimate Std. Error         df t value Pr(>|t|)
```

```
## (Intercept)    3.55926    0.04012 1545.91034  88.717   <2e-16 ***
## wave_c        -0.05030    0.01176 1591.66268  -4.278    2e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##        (Intr)
## wave_c -0.763
## optimizer (nloptwrap) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00603926 (tol = 0.002, component 1)
```

```r
data_tmp <- data.frame(Anxiety.pred = predict(model.b),
                       wave = model.b@frame[["wave_c"]])

ggplot(data = data_tmp, mapping = aes(x = wave, y = Anxiety.pred)) +
  geom_point(col='orange',cex=0.9) +
  stat_smooth(method="lm", formula = y ~ x,cex=0.8) +
  theme_light()
```



3.D.ii 2 Graph the unconditional growth model
iii.Conduct a growth model with the main IV only

```r
table(dat_long$Race)
```

1.Interpret the fixed and random effects

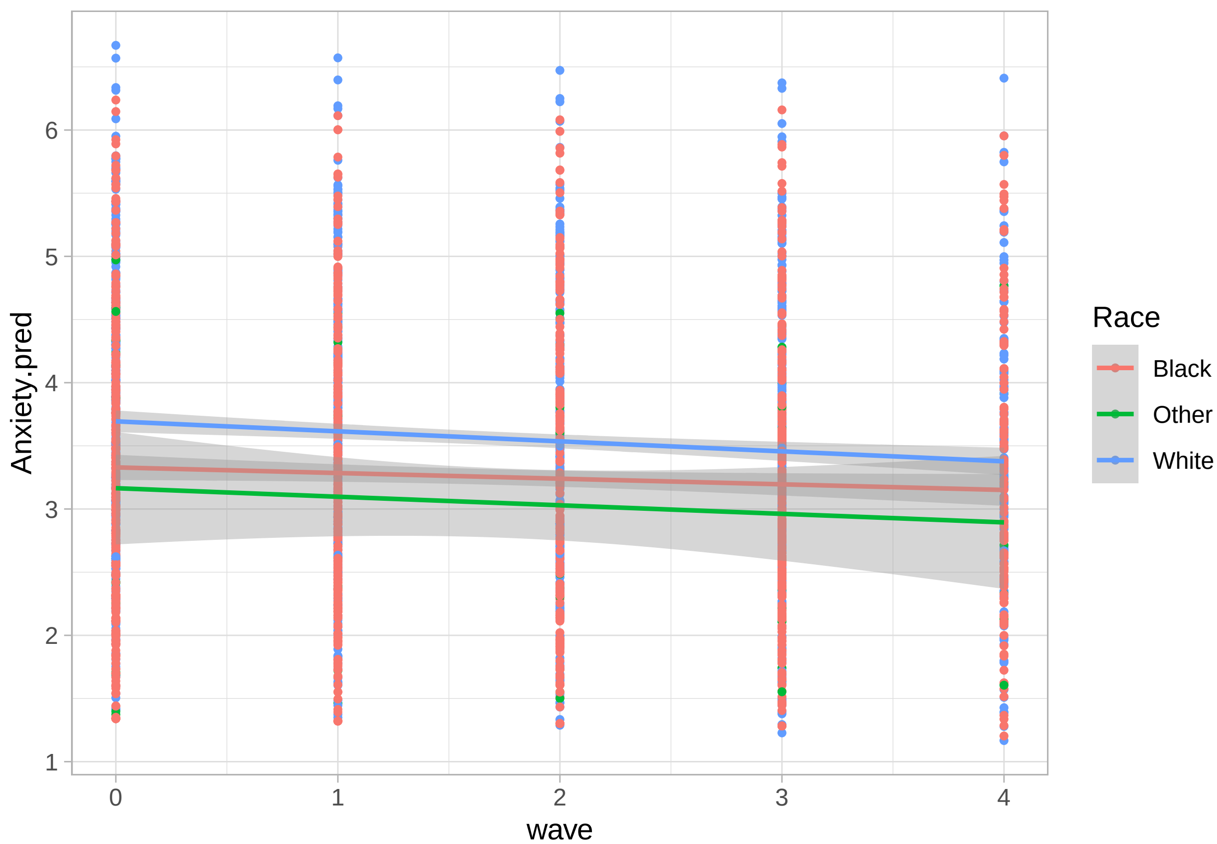
```
##
## Black Other White
##  1560    85  1890
```

```r
model.c <- lmer(Anxiety ~ wave_c*Race + (wave_c|PID), data = dat_long)
summary(model.c)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Anxiety ~ wave_c * Race + (wave_c | PID)
##    Data: dat_long
##
## REML criterion at convergence: 9138.3
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4551 -0.5392 -0.0698  0.5250  4.0960
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  PID      (Intercept) 1.51649  1.2315
##           wave_c      0.05805  0.2409   -0.36
##  Residual             0.85205  0.9231
## Number of obs: 2804, groups:  PID, 707
##
## Fixed effects:
##                    Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)         3.33577    0.08203 702.53580  40.667  < 2e-16 ***
## wave_c             -0.05562    0.02543 636.27986  -2.187  0.02911 *
## RaceOther          -0.19417    0.36296 717.60756  -0.535  0.59284
## RaceWhite           0.35441    0.11079 701.61936   3.199  0.00144 **
## wave_c:RaceOther    0.04833    0.10859 604.70041   0.445  0.65646
## wave_c:RaceWhite   -0.03161    0.03435 631.02573  -0.920  0.35775
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) wave_c RcOthr RacWht wv_:RO
## wave_c       -0.504
## RaceOther    -0.226  0.114
## RaceWhite    -0.740  0.373  0.167
## wv_c:RcOthr   0.118 -0.234 -0.508 -0.087
## wav_c:RcWht   0.373 -0.740 -0.084 -0.503  0.173
```

```r
df.plot.c <- data.frame(Anxiety.pred = predict(model.c),
                        wave = model.c@frame[["wave_c"]],
                        Race = model.c@frame[["Race"]])
```

```
ggplot(data = df.plot.c, mapping = aes(x = wave, y = Anxiety.pred, group = Race, color = Race)) +
  geom_point(cex=0.9) +
  stat_smooth(method="lm", formula = y ~ x,cex=0.8)+theme_light()
```



iv.Conduct a growth model with the main IV and at least one additional time-varying covariate

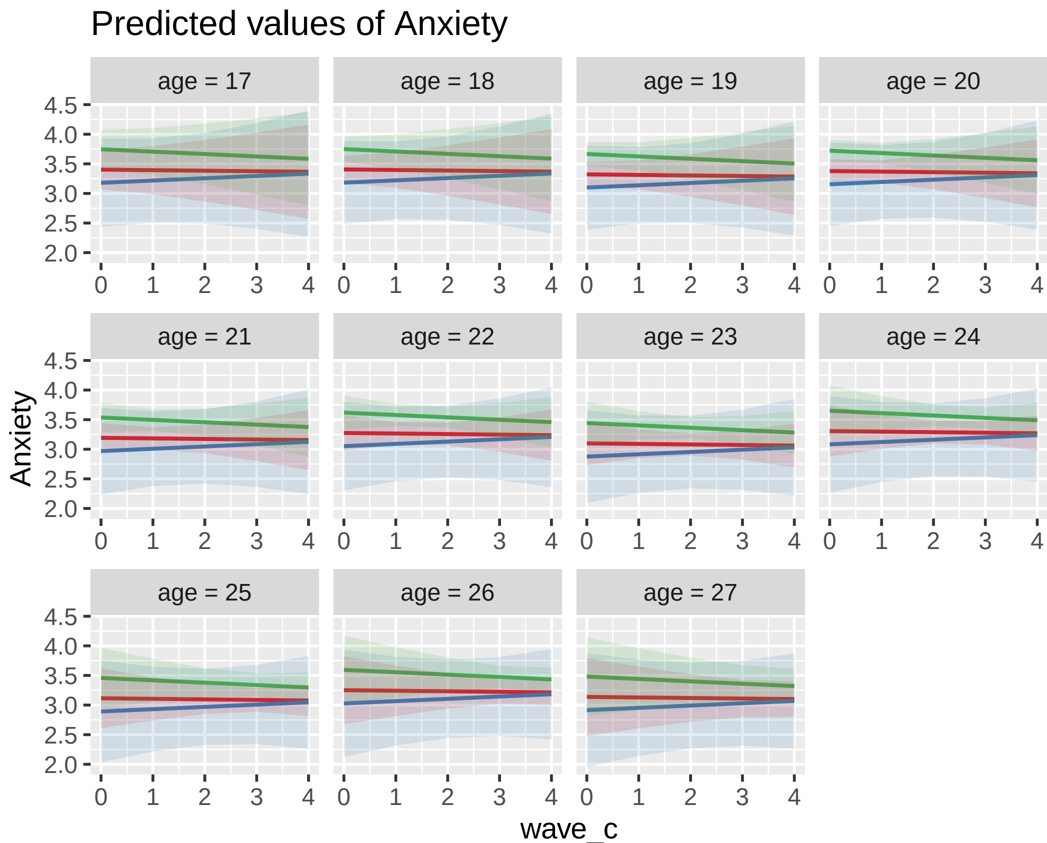
```
model.d <- lmer(Anxiety ~ wave_c*Race + factor(age) + (wave_c|PID), data = dat_long, REML = FALSE)
summary(model.d)
```

1.Interpret the fixed and random effects

```
## Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's
##   method [lmerModLmerTest]
## Formula: Anxiety ~ wave_c * Race + factor(age) + (wave_c | PID)
##    Data: dat_long
##
##      AIC      BIC   logLik deviance df.resid
##   9147.2   9266.0  -4553.6   9107.2     2784
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -3.4795 -0.5371 -0.0654  0.5147  4.1575
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  PID      (Intercept) 1.51572  1.2311
##           wave_c      0.05738  0.2395   -0.37
##  Residual             0.84721  0.9204
## Number of obs: 2804, groups:  PID, 707
##
## Fixed effects:
##                    Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)       3.403e+00  1.701e-01  1.557e+03  20.005  < 2e-16 ***
## wave_c           -9.306e-03  8.138e-02  1.341e+03  -0.114  0.90897
## RaceOther        -2.235e-01  3.637e-01  7.228e+02  -0.614  0.53908
## RaceWhite         3.437e-01  1.109e-01  7.062e+02   3.099  0.00202 **
## factor(age)18     3.070e-03  1.705e-01  1.727e+03   0.018  0.98564
## factor(age)19    -8.012e-02  1.642e-01  2.200e+03  -0.488  0.62572
## factor(age)20    -2.435e-02  1.832e-01  1.741e+03  -0.133  0.89431
## factor(age)21    -2.099e-01  2.039e-01  1.899e+03  -1.029  0.30348
## factor(age)22    -1.272e-01  2.348e-01  1.611e+03  -0.542  0.58816
## factor(age)23    -3.027e-01  2.627e-01  1.590e+03  -1.152  0.24932
## factor(age)24    -9.559e-02  2.978e-01  1.415e+03  -0.321  0.74824
## factor(age)25    -2.876e-01  3.316e-01  1.446e+03  -0.867  0.38592
## factor(age)26    -1.513e-01  3.676e-01  1.315e+03  -0.411  0.68077
## factor(age)27    -2.646e-01  4.054e-01  1.449e+03  -0.653  0.51400
## wave_c:RaceOther  4.816e-02  1.084e-01  6.112e+02   0.444  0.65709
## wave_c:RaceWhite -3.097e-02  3.431e-02  6.358e+02  -0.903  0.36707
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot_model(model.d, type = "pred", terms = c("wave_c", "Race", "age"))
```



2.Graph the growth model

v.Using the fit statistics learned in class (i.e. Likelihood, Devianceand AIC/BIC) assess the model fit between the 4 models conducted. Which is the best model and why?

```
df <- data.frame(fit.stats = c("-2LL", "Deviance", "AIC", "BIC"),
                 model.a = c(-2*logLik(model.a), deviance(model.a), AIC(model.a), BIC(model.a)),
                 model.b = c(-2*logLik(model.b), deviance(model.b), AIC(model.b), BIC(model.b)),
                 model.c = c(-2*logLik(model.c), deviance(model.c), AIC(model.c), BIC(model.c)),
                 model.d = c(-2*logLik(model.d), deviance(model.d), AIC(model.d), BIC(model.d)))
pander(df,caption='Model Comparison')
```

Table 1: Model Comparison

fit.stats	model.a	model.b	model.c	model.d
-2LL	23924	23854	9138	9107
Deviance	23924	23854	9138	9107
AIC	23930	23866	9158	9147
BIC	23950	23907	9218	9266