560 Group Project: Models

Brittany Jones, Tess Abrahamson-Richards, Natalie Turner, Hanyi Wang

2023-03-14

Models fitting

- Model 0: Random Intercepts, no covariates.
- Model 1: Random Intercepts, with main IV.
- Model 2: Random Intercepts, with demographic covariates, shared slopes.
- Model 3: Random Intercepts, with demographic + health status covariates, shared slopes.
- Model 4: Random Intercepts, with main IV, demographic + health status + time.
- Model 5: Random Intercepts, with main IV, demographic + health status + time shared + time^2.
- Model 6: Random Intercepts, with main IV, demographic + health status + time shared + time^2 + varying time.
- Model 7: Random Intercepts, with main IV, demographic + health status + time shared + time^2 + varying time slopes and time^2 slopes.

Models selection based on BIC and AIC

Table 1: AIC and BIC for the 7 models

models	df	AIC	BIC
0	3	152734.4	152764.4
1	4	152601.4	152641.4
2	30	151687.4	151987.5
3	34	148144.3	148484.1
4	35	148152.4	148502.2
5	36	146840.0	147199.8
6	32	146750.8	147070.6
7	41	147200.8	147610.6

Results

Interpretion

```
# # point estimates of shared coefficients
# fixef(M6)

## Within-group variance (residual variance)
sigma(M6)^2
```

```
## [1] 0.1097712
```

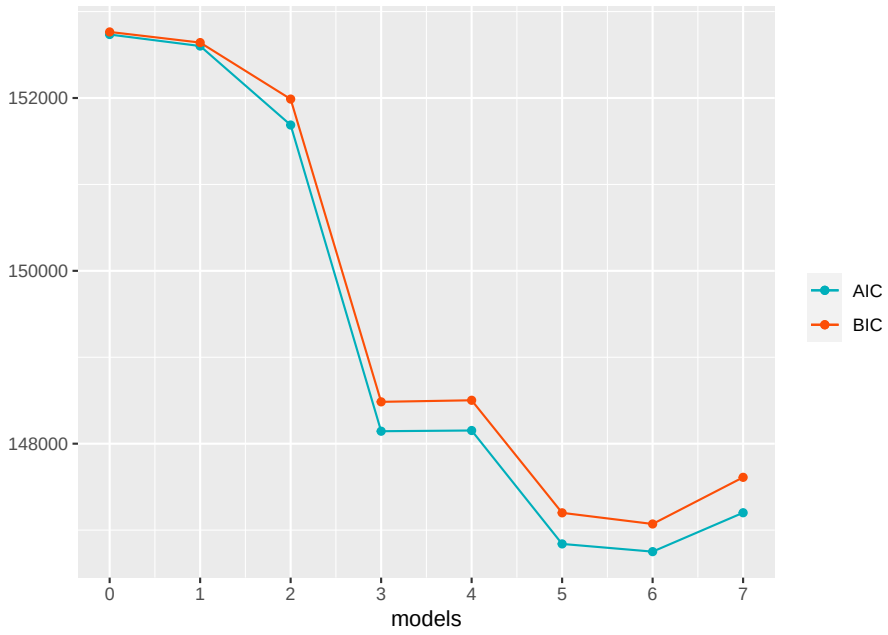


Figure 1: AIC and BIC for the 7 models

```
# between-group variance: 0.4957805^2
VarCorr(M6)
```

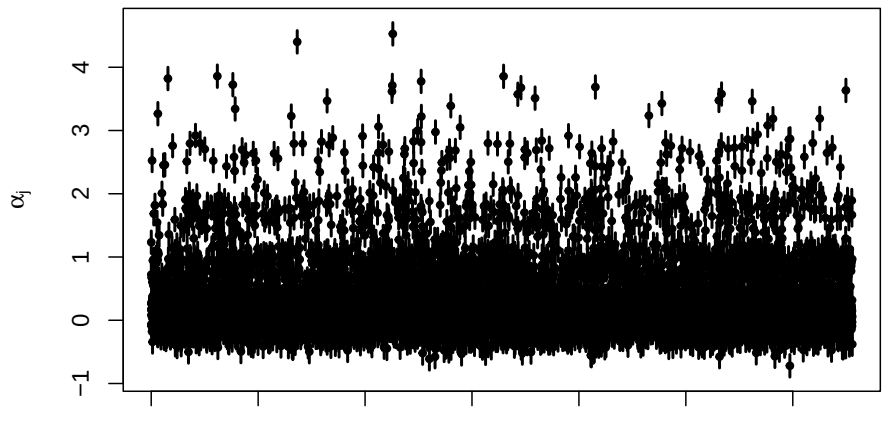
```
##  Groups   Name        Std.Dev.  Corr
##  UQID     (Intercept) 0.4955298
##           MONTH.num   0.0024787 1.000
##  Residual             0.3313174
```

```
## Between-school variance
# variance of sample means
```

p-values

Table 2: p values for covariates of interest

covariates	coefficient	pvalues
log(Health costs)	-0.004545	5.666e-10
Month	0.036410	2.000e+00
Month^2	-0.002855	3.140e-31
Race: Black alone	0.125800	2.000e+00
Race: Residual	0.006292	1.149e+00
Race: White alone	-0.071490	1.384e-03



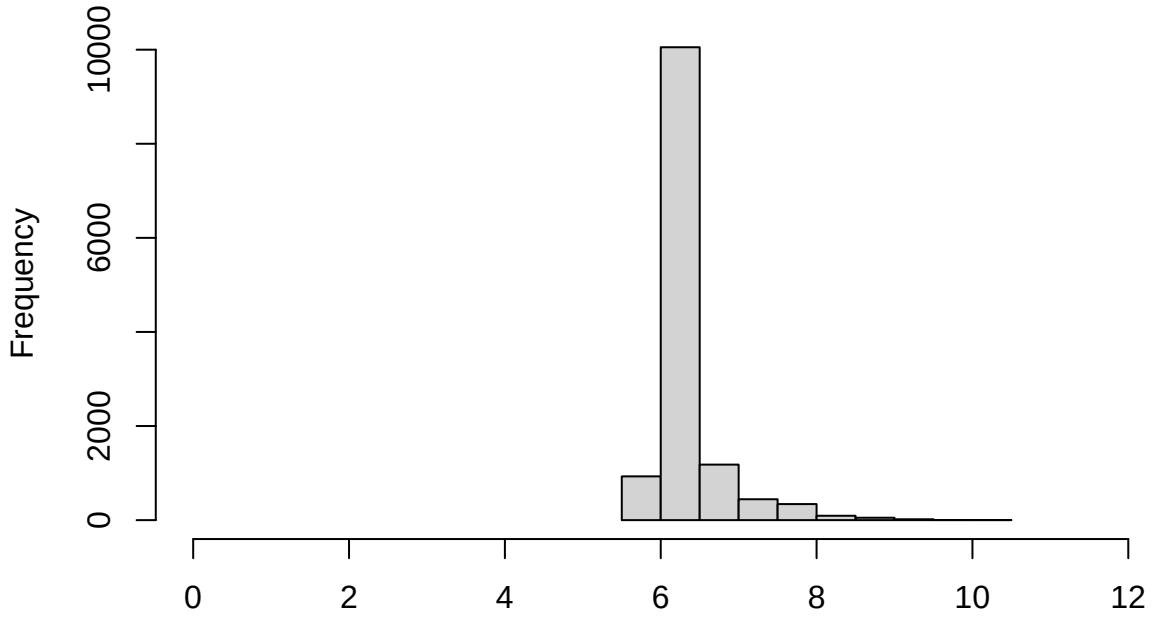


Figure 3: Density of social score utilization minimum points for individuals

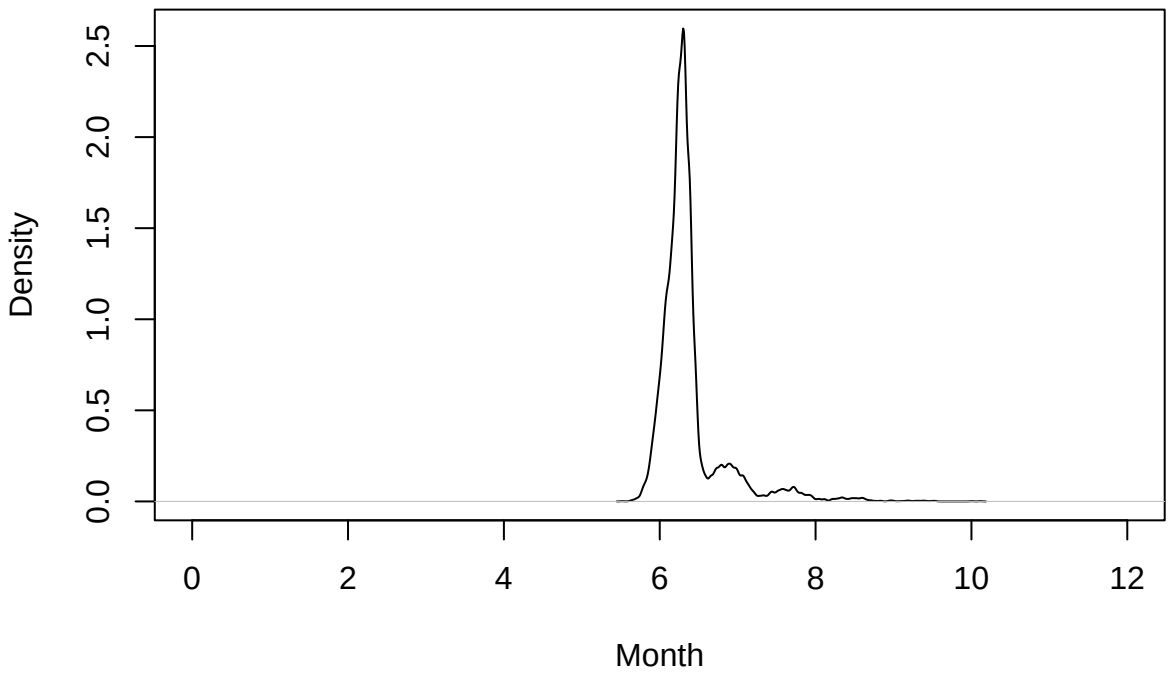


Figure 4: Density of social score utilization minimum points for individuals

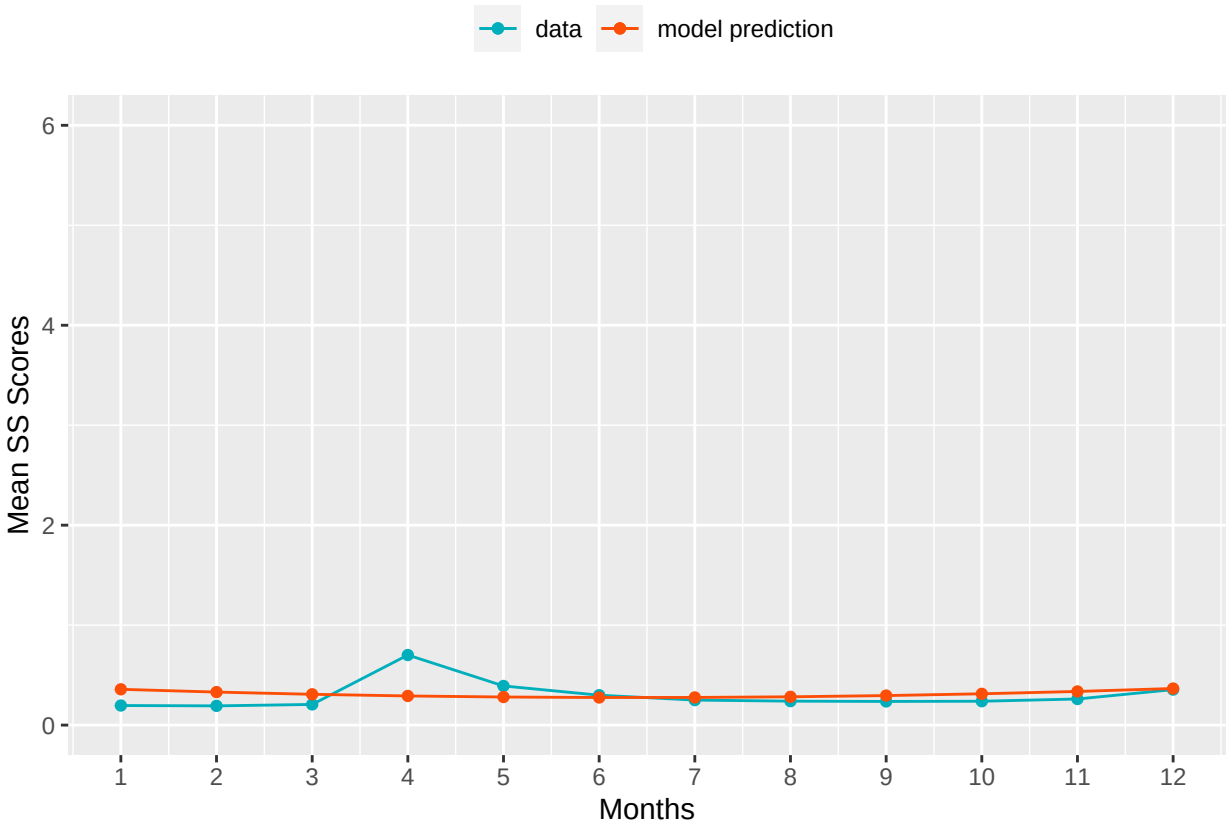


Figure 5: Mean of SS scores for each month based on the data and model prediction

Code Appendix

```
knitr::opts_chunk$set(echo = FALSE,message = FALSE,warning = FALSE)
library(tidyverse)
library(mlmRev)
library(lme4)
library(arm)
library(haven)
library(table1)
library(Hmisc)
library(ggplot2)
library(ggpubr)
library(expss)
## ------------- working directories for Hanyi ----------
wd_hanyi = '/Users/hanyiwang/Desktop/Social-Service-Participation-Among-Older-Adult-Women-in-Washington-
path_hanyi = c("../SIPPdata/sipp.csv")

## ------------- read data ----------
setwd(wd_hanyi)
sipp = read.csv(path_hanyi)

## ------------- data cleaning ----------
sipp = sipp %>%
  mutate(MONTH.num = as.numeric(MONTHCODE)) %>%
  mutate(UQID = factor(UQID)) %>%
  mutate(HI.log = log(HI + .0001)) # add 0.0001 to avoid 0
# Model 0: Random Intercepts, no covariates
M0 <- lmer(SS_SCORE ~ 1 + (1 | UQID), data = sipp)

# Model 1: Random Intercepts, with main IV
M1 <- lmer(SS_SCORE ~ 1 + (HI.log) + (1 | UQID), data = sipp)

# Model 2: Random Intercepts, with demographic covariates, shared slopes
M2 <- lmer(SS_SCORE ~ 1 + (HI.log) + TAGE + TRACE + EORIGIN + EEDUC + (1 | UQID), data = sipp)

# Model 3: Random Intercepts, with demographic + health status covariates, shared slopes
M3 <- lmer(SS_SCORE ~ 1 + (HI.log) + TAGE + TRACE + EORIGIN + EEDUC + EHLTSTAT + (1 | UQID), data = sip

# Model 4: Random Intercepts, with main IV, demographic + health status + time
M4 <- lmer(SS_SCORE ~ 1 + (HI.log) + TAGE + TRACE + EORIGIN + EEDUC + EHLTSTAT +
             MONTH.num + (1 | UQID), data = sipp)

# Model 5: Random Intercepts, with main IV, demographic + health status + time shared + time^2
M5 <- lmer(SS_SCORE ~ 1 + (HI.log) + TAGE + TRACE + EORIGIN + EEDUC + EHLTSTAT +
             MONTH.num + I(MONTH.num^2) + (1 | UQID), data = sipp)

# Model 6: Random Intercepts, with main IV, demographic + health status + time shared + time^2 + varyin
M6 <- lmer(SS_SCORE ~ 1 + (HI.log) + TAGE + ERACE + EORIGIN + EEDUC + EHLTSTAT +
             MONTH.num + I(MONTH.num^2) + (1 + MONTH.num | UQID), data = sipp)

# Model 7: Random Intercepts, with main IV, demographic + health status + time shared + time^2 + varyin
M7 <- lmer(SS_SCORE ~ 1 + (HI.log) + TAGE + TRACE + EORIGIN + EEDUC + EHLTSTAT +
           MONTH.num + I(MONTH.num^2) + (1 + MONTH.num +I(MONTH.num^2)| UQID), data = sipp)
aic.bic = data.frame(
```

```
  models = 0:7,
  df = AIC(M0, M1, M2, M3, M4, M5, M6,M7)$df,
  AIC = AIC(M0, M1, M2, M3, M4, M5, M6,M7)$AIC,
  BIC = BIC(M0, M1, M2, M3, M4, M5, M6,M7)$BIC)

knitr::kable(aic.bic,caption = 'AIC and BIC for the 7 models')

ggplot(aic.bic,aes(x = models,y=AIC)) +
  geom_point(aes(x = models,y=AIC,color = "AIC")) +
  geom_line(aes(x = models,y=AIC,color = "AIC")) +
  geom_point(aes(x = models,y=BIC,color = "BIC")) +
  geom_line(aes(x = models,y=BIC,color = "BIC")) +
  labs(x = "models",y = "") +
  scale_colour_manual(name="",values=c("AIC"="#00AFBB","BIC"="#FC4E07")) +
  scale_x_continuous(breaks=seq(0,7,by=1))

# # point estimates of shared coefficients
# fixef(M6)

## Within-group variance (residual variance)
sigma(M6)^2

# between-group variance: 0.4957805^2
VarCorr(M6)

## Between-school variance
# variance of sample means

# summary(M6) # find t values
pvalues = data.frame(
  covariates = c("log(Health costs)","Month","Month^2",
                 "Race: Black alone","Race: Residual","Race: White alone"),
  coefficient = c(-4.545e-03,3.641e-02,-2.855e-03,1.258e-01,6.292e-03,-7.149e-02 ),
  pvalues = c(2*pt(-8.232,df=39-1) ,2*pt(34.921,df=39-1), 2*pt(-36.579,df=39-1),
              2*pt(5.224,df=39-1) ,2*pt(0.189,df=39-1) ,2*pt(-3.451,df=39-1)))

pvalues$pvalues = formatC(pvalues$pvalues, format = "e", digits = 3)
knitr::kable(pvalues,caption = 'p values for covariates of interest')

# fixef(M6)

ptEstAndCIs <- data.frame(
  coef(M6)$UQID[,"(Intercept)"],
  coef(M6)$UQID[,"(Intercept)"]-se.ranef(M6)$UQID[,"(Intercept)"]*1.96,
  coef(M6)$UQID[,"(Intercept)"]+se.ranef(M6)$UQID[,"(Intercept)"]*1.96
)

colnames(ptEstAndCIs) <- c("RE_ests","RE_L","RE_U")

plot(x = 1:nrow(ptEstAndCIs),
     y = ptEstAndCIs$RE_ests,
     ylim = range(ptEstAndCIs$RE_L, ptEstAndCIs$RE_U),
     xlab = "Participants",
```

```
     ylab = expression(paste(alpha[j])),
     pch = 20)
segments(x0 = 1:nrow(ptEstAndCIs),
         y0 = ptEstAndCIs$RE_L, y1 = ptEstAndCIs$RE_U, lwd = 2)

axis.sym.vec = -(coef(M6)$UQID[27])/(2*coef(M6)$UQID[28])
axis.sym.vec = data.frame(peak = array(axis.sym.vec[[1]]))

hist(axis.sym.vec$peak,xlim=c(0,12),xlab = "",main = "")
plot(density(axis.sym.vec$peak),xlim=c(0,12),xlab = "Month",main = "")

months.mean = c()
for (i in 1:12){months.mean[i] = sipp %>% filter(MONTHCODE==i) %>% dplyr::select(SS_SCORE) %>% mean() }

# sipp %>% dplyr::select(HI) %>% mean()
# sipp %>% dplyr::select(TAGE) %>% mean()
# summary(M6)[["coefficients"]]
month = c(1,2,3,4,5,6,7,8,9,10,11,12)
model.pred = 0.258358491 - 0.004544700 * (log(1230)) - 0.002060997*(67) -0.071489698 - 0.050425209 + 0.1

ss.mean = data.frame(
  months = 1:12,
  data = months.mean,
  model = model.pred)

ggplot(ss.mean,aes(x = months,y=data)) +
  geom_point(aes(x = months,y=data,color = "data")) +
  geom_line(aes(x = months,y=data,color = "data")) +
  geom_point(aes(x = months,y=model,color = "model prediction")) +
  geom_line(aes(x = months,y=model,color = "model prediction")) +
  labs(x = "Months",y = "Mean SS Scores") +
  scale_colour_manual(name="",values=c("data"="#00AFBB","model prediction"="#FC4E07")) +
  xlim(0,12)+scale_x_continuous(breaks=seq(0,12,by=1))+
  ylim(0,6)+ theme(legend.position = "top")

# plot(months.mean,type = "o",xlim=c(0,12),xlab = "Month",ylab = "SS Score",main = "")
```