Longitudinal_Gee

Haotian Zheng

2025-02-08

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(mice)
```

```
##
## Attaching package: 'mice'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     cbind, rbind
```

```
library(gee)
library(gtsummary)
library(emmeans)
```

```
## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
```

```
library(ggeffects)
library(geepack)
library(table1)
```

```
##
## Attaching package: 'table1'
##
## The following objects are masked from 'package:base':
##
##     units, units<-
```

```
'https://rpubs.com/izf381'
```

```
## [1] "https://rpubs.com/izf381"
```

```
library(tidyverse)
# Loading data
load("/Users/haotian/Documents/Add_Health_Dataset/21600-0001-Data.rda")
load("/Users/haotian/Documents/Add_Health_Dataset/21600-0022-Data.rda")
load("/Users/haotian/Documents/Add_Health_Dataset/21600-0036-Data.rda")
load("/Users/haotian/Documents/Add_Health_Dataset/21600-0032-Data.rda")
load("/Users/haotian/Documents/Add_Health_Dataset/21600-0042-Data.rda")

w1 = da21600.0001 %>% dplyr::select(H1GI1Y,AID,BIO_SEX,H1GI4,H1GI6A,H1GI6B,H1GI6C,H1GI6D,H1GI6E,H1GH59A
w2 = da21600.0022
w3 = da21600.0032

# weights
ww = da21600.0042 %>% dplyr::select(AID,GSW145)

plote = function(x){
  emmeans(x, specs = c("wave", "autonomy"),
          at = list(diagnose = "obs"),
          cov.keep = "wave",
          regrid = "response") %>%
    as.data.frame() %>% ggplot() +
    aes(x = wave, y = prob, color = autonomy, fill = autonomy) +
    geom_line() +
    geom_ribbon(aes(ymin = asymp.LCL, ymax = asymp.UCL,
                    color = NULL), alpha = 0.15) +
    scale_color_brewer(palette = "Set1") +
    scale_fill_brewer(palette = "Set1", guide = NULL) +
    scale_y_continuous(labels = scales::percent) +
    theme_ggeffects() +
    labs(title = "GEE Effect plot", y = "obesity")
}
```

Cleaning & Recoding

```
w1$H1GI6A = as.numeric(w1$H1GI6A)- 1
w1$H1GI6B = as.numeric(w1$H1GI6B)- 1
w1$H1GI6C = as.numeric(w1$H1GI6C)- 1
w1$H1GI6D = as.numeric(w1$H1GI6D)- 1
w1$H1GI6E = as.numeric(w1$H1GI6E)- 1
```

```
w1$w1age = 21 - as.integer(w1$H1GI1Y)
w1$racecount = w1$H1GI6A+w1$H1GI6B+w1$H1GI6C+w1$H1GI6D+w1$H1GI6E
w1$RACE = ifelse(as.numeric(w1$H1GI6A) == "1","White",NA)
w1$RACE = ifelse(as.numeric(w1$H1GI6B) == "1","Black",w1$RACE)
w1$RACE = ifelse(as.numeric(w1$H1GI6C) == "1","Indian",w1$RACE)
w1$RACE = ifelse(as.numeric(w1$H1GI6D) == "1","Asian",w1$RACE)
w1$RACE = ifelse(as.numeric(w1$H1GI6E) == "1","Other",w1$RACE)
w1$RACE = ifelse(as.numeric(w1$racecount) > 1,"Multi-Race",w1$RACE)
w1$RACE = ifelse(as.numeric(w1$H1GI4) == "2","Hispanic",w1$RACE)
w1$PAEDUC = ifelse(as.numeric(w1$PA12) < 4,"Less than High School",NA)
w1$PAEDUC = ifelse(as.numeric(w1$PA12) %in% c(4,5),"High School Graduate",w1$PAEDUC)
w1$PAEDUC = ifelse(as.numeric(w1$PA12) %in% c(6,7),"Some College",w1$PAEDUC)
w1$PAEDUC = ifelse(as.numeric(w1$PA12) == 8 ,"College Graduate",w1$PAEDUC)
w1$PAEDUC = ifelse(as.numeric(w1$PA12) == 9 ,"College Graduate +",w1$PAEDUC)
w1$PAEDUC = ifelse(as.numeric(w1$PA12) == 10 ,"Less than High School",w1$PAEDUC)
w1$EDUC1 = "Less than High School"
w1$w1HI = ifelse(w1$PA55<75,"Less than $75000",NA)
w1$w1HI = ifelse(w1$PA55<100 & w1$PA55>75,"$75000 - $99999",w1$w1HI)
w1$w1HI = ifelse(w1$PA55>100,"$100000 and over",w1$w1HI)




w2$w2HI = ifelse(as.numeric(w2$H4EC1)<10,"Less than $75000",NA)
w2$w2HI = ifelse(as.numeric(w2$H4EC1) == 10,"$75000 - $99999",w2$w2HI)
w2$w2HI = ifelse(as.numeric(w2$H4EC1) %in% c(11,12),"$100000 and over",w2$w2HI)
w2$EDUC2 = ifelse(as.numeric(w2$H4ED2)<3,"Less than High School",NA)
w2$EDUC2 = ifelse(as.numeric(w2$H4ED2)==3,"High School Graduate",w2$EDUC2)
w2$EDUC2 = ifelse(as.numeric(w2$H4ED2)%in%c(4,5,6),"Some College",w2$EDUC2)
w2$EDUC2 = ifelse(as.numeric(w2$H4ED2) == 7,"College Graduate",w2$EDUC2)
w2$EDUC2 = ifelse(as.numeric(w2$H4ED2) %in% c(8,9,10,11,12,13),"College Graduate +",w2$EDUC2)
w2$H4DA5 = as.numeric(w2$H4DA5)




w3$w3HI = ifelse(as.numeric(w3$H5EC2)<10,"Less than $75000",NA)
w3$w3HI = ifelse(as.numeric(w3$H5EC2) == 10,"$75000 - $99999",w3$w3HI)
w3$w3HI = ifelse(as.numeric(w3$H5EC2) %in% c(11,12),"$100000 and over",w3$w3HI)
w3$EDUC3 = ifelse(as.numeric(w3$H5OD11) == 2, "Less than High School",NA)
w3$EDUC3 = ifelse(as.numeric(w3$H5OD11) %in% c(3,4), "Less than High School",w3$EDUC3)
w3$EDUC3 = ifelse(as.numeric(w3$H5OD11) %in% c(5,6,7,8,9), "Some College",w3$EDUC3)
w3$EDUC3 = ifelse(as.numeric(w3$H5OD11) ==10, "College Graduate",w3$EDUC3)
w3$EDUC3 = ifelse(as.numeric(w3$H5OD11) %in% c(11,12,13,14,15,16), "College Graduate +",w3$EDUC3)
w3$H5ID27 = as.numeric(w3$H5ID27)



ready = ww %>% merge(w3,by="AID") %>% merge(w2,by = "AID") %>% merge(w1,by ="AID")  %>%
  dplyr::select(AID,H1GH60,H1GH59A,H1GH59B,H4GH5F,H4GH5I,H4GH6,H5ID2F,H5ID2I,H5ID3,PAEDUC,RACE,BIO_SEX,
```

```
ready$AID = as.numeric(ready$AID)


ready$w2age = ready$w1age+14
ready$w3age = ready$w2age + 8


ready$H4DA5 <- as.numeric(ready$H4DA5)
ready$H5ID27 <- as.numeric(ready$H5ID27)

str(ready)
```

```
## 'data.frame':    3713 obs. of  26 variables:
##  $ AID    : num  1 2 4 6 7 8 9 11 13 14 ...
##  $ H1GH60 : num  152 110 280 75 130 105 108 150 205 83 ...
##  $ H1GH59A: Factor w/ 3 levels "(4) (4) 4 feet",..: 3 2 3 NA 2 2 2 2 2 2 ...
##  $ H1GH59B: Factor w/ 12 levels "(00) (0) 0 inches",..: 1 1 3 NA 6 3 1 7 12 2 ...
##  $ H4GH5F : num  5 5 6 5 5 5 5 5 6 5 ...
##  $ H4GH5I : num  11 9 2 6 5 2 3 10 1 4 ...
##  $ H4GH6  : num  238 200 325 125 150 139 140 200 294 120 ...
##  $ H5ID2F : num  5 5 6 5 5 5 5 5 6 5 ...
##  $ H5ID2I : num  11 9 2 8 5 2 3 10 3 4 ...
##  $ H5ID3  : num  275 210 340 135 160 130 NA 240 255 120 ...
##  $ PAEDUC : chr  NA "Some College" "Some College" "College Graduate" ...
##  $ RACE   : chr  "Black" "White" "Black" "White" ...
##  $ BIO_SEX: Factor w/ 2 levels "(1) (1) Male",..: 2 1 1 2 2 2 2 1 1 2 ...
##  $ EDUC1  : chr  "Less than High School" "Less than High School" "Less than High School" "Less than H
##  $ EDUC2  : chr  "High School Graduate" "Some College" "Some College" "Some College" ...
##  $ EDUC3  : chr  "Less than High School" "Some College" "Some College" "Some College" ...
##  $ H4DA5  : num  0 0 0 0 0 2 0 0 0 0 ...
##  $ H5ID27 : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ H1WP7  : Factor w/ 2 levels "(0) (0) No","(1) (1) Yes": 2 1 2 2 2 2 2 1 1 2 ...
##  $ w1HI   : chr  NA "Less than $75000" "Less than $75000" "Less than $75000" ...
##  $ w2HI   : chr  "Less than $75000" "Less than $75000" "Less than $75000" "Less than $75000" ...
##  $ w3HI   : chr  "Less than $75000" "Less than $75000" "Less than $75000" "Less than $75000" ...
##  $ GSW145 : num  666 1057 870 1831 941 ...
##  $ w1age  : num  18 13 15 12 15 15 14 13 13 12 ...
##  $ w2age  : num  32 27 29 26 29 29 28 27 27 26 ...
##  $ w3age  : num  40 35 37 34 37 37 36 35 35 34 ...
```

```
ready$w1HI = as.factor(ready$w1HI)
ready$w2HI=as.factor(ready$w2HI)
ready$w3HI=as.factor(ready$w3HI)
ready$EDUC1=as.factor(ready$EDUC1)
ready$EDUC2=as.factor(ready$EDUC2
                      )
ready$EDUC3=as.factor(ready$EDUC3)
ready$PAEDUC = as.factor(ready$PAEDUC)
ready$RACE = as.factor(ready$RACE)
levels(ready$w1HI) = c("Less than $75000","$75000 - $99999","$100000 and over")
levels(ready$w2HI)= c("Less than $75000","$75000 - $99999","$100000 and over")
levels(ready$w3HI)= c("Less than $75000","$75000 - $99999","$100000 and over")
```

```
levels(ready$PAEDUC) = c("Less than High School", "High School Graduate","Some College","College Graduat
levels(ready$EDUC1) = c("Less than High School", "High School Graduate","Some College","College Graduate
levels(ready$EDUC2) = c("Less than High School", "High School Graduate","Some College","College Graduate
levels(ready$EDUC3) = c("Less than High School","Some College","College Graduate","College Graduate +")
```

```
ready$BIO_SEX = as.factor(ifelse(ready$BIO_SEX == '(1) (1) Male',"Male","Famale"))
colnames(ready)[17] = "w2spt"
colnames(ready)[18] = 'w3spt'
colnames(ready)[19] = 'autonomy'
ready$autonomy = as.factor(ifelse(ready$autonomy == "(0) (0) No","No","Yes"))
ready$H1GH59A = as.numeric(ready$H1GH59A)+3

ready$H1GH59B = as.numeric(ready$H1GH59B)-1
ready$H5ID2F = ifelse(ready$H5ID2F > 8,NA,ready$H5ID2F)
ready$H5ID2I = ifelse(ready$H5ID2I > 12,NA,ready$H5ID2I)
```

Imputation & Format Change

```
library(mice)
ini = mice(ready,seed = 1,m = 3)
```

```
##
##  iter imp variable
##    1   1  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    1   2  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    1   3  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    2   1  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    2   2  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    2   3  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    3   1  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    3   2  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    3   3  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    4   1  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    4   2  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    4   3  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    5   1  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    5   2  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    5   3  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
```

```
## Warning: Number of logged events: 3
```

```
pred1 = ini$predictorMatrix
pred1[,'AID'] = 0
ready = complete(mice(ready,seed = 1,pred = pred1,m = 3))
```

```
##
```

```
##  iter imp variable
##    1   1  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    1   2  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    1   3  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    2   1  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    2   2  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    2   3  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    3   1  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    3   2  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    3   3  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    4   1  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    4   2  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    4   3  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    5   1  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    5   2  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
##    5   3  H1GH60  H1GH59A  H1GH59B  H4GH5F  H4GH5I  H4GH6  H5ID2F  H5ID2I  H5ID3  PAEDUC  RACE  EDUC2
```

```
# Construction of BMI & Obesity

ready$w1obs =



  (703*(ready$H1GH60)/

  ((
    ((as.integer(ready$H1GH59A))*12) + ((as.numeric(ready$H1GH59B))))^2))


ready$w2obs =



  (703*(ready$H4GH6)/

     ((
       ((as.integer(ready$H4GH5F))*12) + ((as.numeric(ready$H4GH5I))))^2))

ready$w3obs =



  (703*(ready$H5ID3)/

     ((
       ((as.integer(ready$H5ID2F))*12) + ((as.numeric(ready$H5ID2I))))^2))

ready %>% filter(w2obs <= 10)
```

```
##    AID H1GH60 H1GH59A H1GH59B H4GH5F H4GH5I H4GH6 H5ID2F H5ID2I H5ID3
## 1  727    175       5       8      5      9    20      5      9   220
## 2 1594    145       5       3      5      4    22      5      4   280
## 3 3705    120       5       2      5      5    22      5      6   225
## 4 4195    145       5       9      6      1    22      5      1   230
```

```
##                 PAEDUC  RACE BIO_SEX                 EDUC1
## 1         Some College White    Male Less than High School
## 2         Some College White  Famale Less than High School
## 3   College Graduate + White  Famale Less than High School
## 4 High School Graduate Black    Male Less than High School
##                   EDUC2                EDUC3 w2spt w3spt autonomy
## 1  High School Graduate          Some College     0     0      Yes
## 2          Some College      College Graduate     0     0      Yes
## 3    College Graduate +    College Graduate +     0     0       No
## 4 Less than High School Less than High School     0     0      Yes
##               w1HI             w2HI             w3HI    GSW145 w1age w2age
## 1  $75000 - $99999 $100000 and over Less than $75000  579.4677    18    32
## 2 $100000 and over $100000 and over $100000 and over 1297.2217    15    29
## 3 $100000 and over Less than $75000  $75000 - $99999 1062.3162    12    26
## 4 $100000 and over $100000 and over Less than $75000 1603.6172    13    27
##   w3age    w1obs    w2obs    w3obs
## 1    40 26.60575 2.953161 32.48477
## 2    37 25.68279 3.775879 48.05664
## 3    34 21.94589 3.660592 36.31198
## 4    35 21.41042 2.902233 43.45337
```

```
# Removing duplicates & unnecesarries

ready = ready %>% dplyr::select(-H1GH60,-H1GH59A,-H1GH59B,-H4GH5F,-H4GH5I,-H4GH6,-H5ID2F,-H5ID2I,-H5ID3

# Format Change



ready$baselineHI = ready$w1HI
tbl1 = ready
table1(~w1obs+w2obs+w3obs|autonomy+baselineHI,data = tbl1)
```

	No			Yes		
	Less than $75000	$75000 - $99999	$100000 and over	Less than $75000	$75000 - $99999	$
	(N=38)	(N=37)	(N=576)	(N=157)	(N=227)	(
w1obs						
Mean (SD)	20.2 (3.05)	21.4 (4.22)	22.3 (4.52)	21.5 (3.28)	21.8 (3.41)	2
Median [Min, Max]	20.2 [13.2, 27.9]	20.8 [14.8, 36.6]	21.3 [13.2, 43.2]	20.7 [15.7, 32.9]	21.3 [15.1, 33.7]	2
w2obs						
Mean (SD)	25.2 (5.27)	25.7 (5.37)	28.7 (7.20)	25.7 (5.39)	26.3 (6.00)	2
Median [Min, Max]	24.1 [18.3, 37.2]	24.3 [18.4, 39.5]	27.4 [3.66, 57.6]	24.4 [18.2, 51.2]	25.1 [2.95, 48.3]	2
w3obs						
Mean (SD)	26.7 (5.95)	27.0 (5.68)	30.1 (7.42)	26.4 (5.73)	28.0 (6.92)	3
Median [Min, Max]	24.4 [18.8, 42.6]	25.8 [18.0, 41.6]	28.7 [17.2, 62.3]	25.1 [16.0, 48.9]	26.5 [17.8, 58.2]	2

```
colnames(tbl1)[2] = "Parent's Education Level"
colnames(tbl1)[5] = "Participants Education Wave I"
colnames(tbl1)[6] = "Participants Education Wave IV"
colnames(tbl1)[7] = "Participants Education Wave V"
colnames(tbl1)[8] = "Participants Sports Level Wave IV"
```

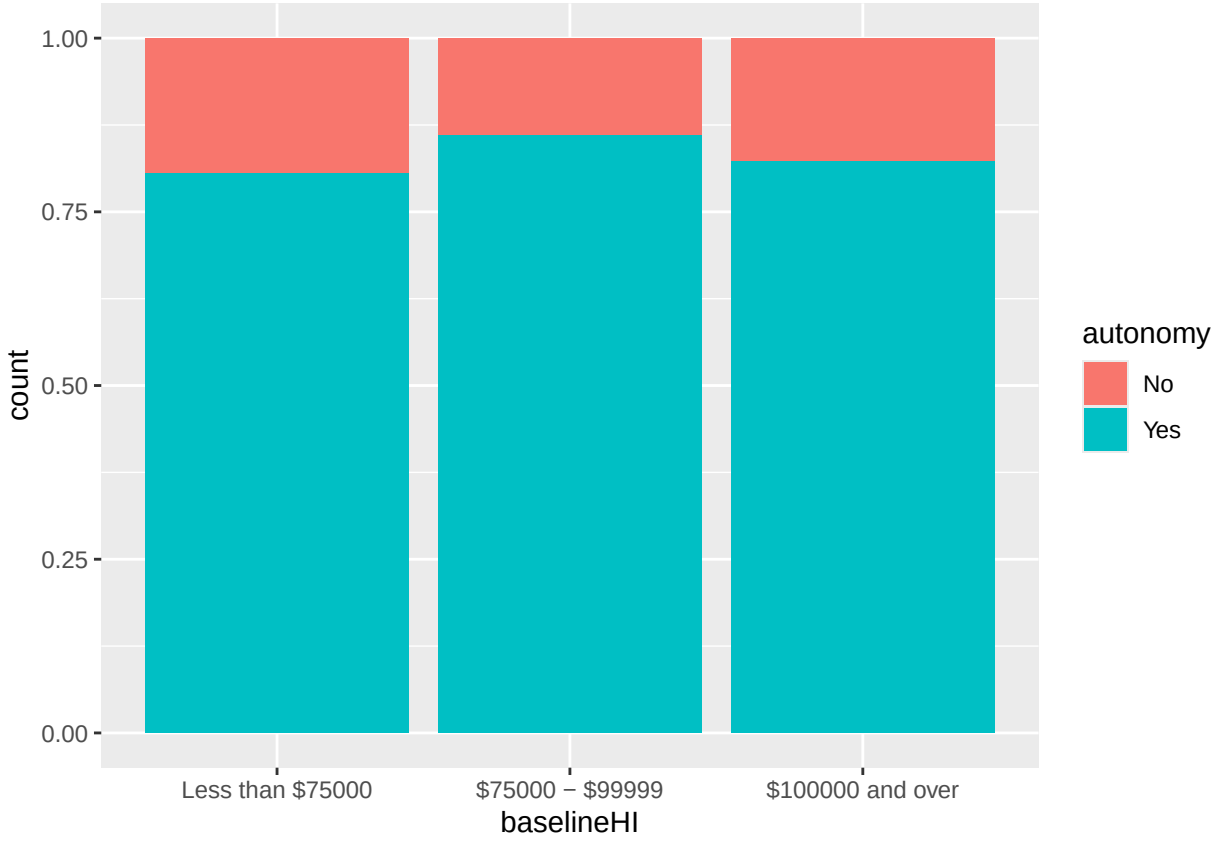
```r
colnames(tbl1)[9] = "Participants Sports Level Wave V"
colnames(tbl1)[11] = "Household Income Wave I"
colnames(tbl1)[12] = "Household Income Wave IV"
colnames(tbl1)[13] = "Household Income Wave V"
colnames(tbl1)[17] = "BMI Wave I"
colnames(tbl1)[18] = "BMI Wave IV"
colnames(tbl1)[19] = "BMI Wave V"
colnames(tbl1)[14] = "Age Wave I"
colnames(tbl1)[15] = "Age Wave IV"
colnames(tbl1)[16] = "Age Wave V"
colnames(tbl1)[4] = "Sex"
colnames(tbl1)[3] = "Race"
colnames(tbl1)[10] = "Autonomy"

table1(~.-AID|baselineHI,data = tbl1)
```

	Less than $75000	$75000 - $99999	$100000 and over	Overall
	(N=195)	(N=264)	(N=3254)	(N=3713)
Parent's Education Level				
Less than High School	53 (27.2%)	78 (29.5%)	456 (14.0%)	587 (15.8%)
High School Graduate	85 (43.6%)	69 (26.1%)	277 (8.5%)	431 (11.6%)
Some College	17 (8.7%)	41 (15.5%)	1042 (32.0%)	1100 (29.6%)
College Graduate	7 (3.6%)	8 (3.0%)	493 (15.2%)	508 (13.7%)
College Graduate +	33 (16.9%)	68 (25.8%)	986 (30.3%)	1087 (29.3%)
Race				
Asian	9 (4.6%)	5 (1.9%)	82 (2.5%)	96 (2.6%)
Black	13 (6.7%)	42 (15.9%)	661 (20.3%)	716 (19.3%)
Hispanic	10 (5.1%)	12 (4.5%)	353 (10.8%)	375 (10.1%)
Indian	0 (0%)	0 (0%)	18 (0.6%)	18 (0.5%)
Multi-Race	8 (4.1%)	7 (2.7%)	131 (4.0%)	146 (3.9%)
Other	0 (0%)	3 (1.1%)	23 (0.7%)	26 (0.7%)
White	155 (79.5%)	195 (73.9%)	1986 (61.0%)	2336 (62.9%)
Sex				
Famale	126 (64.6%)	144 (54.5%)	1879 (57.7%)	2149 (57.9%)
Male	69 (35.4%)	120 (45.5%)	1375 (42.3%)	1564 (42.1%)
Participants Education Wave I				
Less than High School	195 (100%)	264 (100%)	3254 (100%)	3713 (100%)
High School Graduate	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Some College	0 (0%)	0 (0%)	0 (0%)	0 (0%)
College Graduate	0 (0%)	0 (0%)	0 (0%)	0 (0%)
College Graduate +	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Participants Education Wave IV				
Less than High School	68 (34.9%)	95 (36.0%)	653 (20.1%)	816 (22.0%)
High School Graduate	75 (38.5%)	82 (31.1%)	385 (11.8%)	542 (14.6%)
Some College	13 (6.7%)	14 (5.3%)	528 (16.2%)	555 (14.9%)
College Graduate	2 (1.0%)	5 (1.9%)	233 (7.2%)	240 (6.5%)
College Graduate +	37 (19.0%)	68 (25.8%)	1455 (44.7%)	1560 (42.0%)
Participants Education Wave V				
Less than High School	59 (30.3%)	88 (33.3%)	644 (19.8%)	791 (21.3%)
Some College	95 (48.7%)	98 (37.1%)	559 (17.2%)	752 (20.3%)
College Graduate	5 (2.6%)	15 (5.7%)	683 (21.0%)	703 (18.9%)
College Graduate +	36 (18.5%)	63 (23.9%)	1368 (42.0%)	1467 (39.5%)
Participants Sports Level Wave IV				
Mean (SD)	1.05 (1.65)	0.913 (1.68)	0.663 (1.45)	0.701 (1.48)
Median [Min, Max]	0 [0, 7.00]	0 [0, 7.00]	0 [0, 7.00]	0 [0, 7.00]
Participants Sports Level Wave V				
Mean (SD)	0.877 (1.56)	0.761 (1.49)	0.632 (1.42)	0.654 (1.44)
Median [Min, Max]	0 [0, 7.00]	0 [0, 7.00]	0 [0, 7.00]	0 [0, 7.00]
Autonomy				
No	38 (19.5%)	37 (14.0%)	576 (17.7%)	651 (17.5%)
Yes	157 (80.5%)	227 (86.0%)	2678 (82.3%)	3062 (82.5%)
Household Income Wave I				
Less than $75000	195 (100%)	0 (0%)	0 (0%)	195 (5.3%)
$75000 - $99999	0 (0%)	264 (100%)	0 (0%)	264 (7.1%)
$100000 and over	0 (0%)	0 (0%)	3254 (100%)	3254 (87.6%)
Household Income Wave IV				
Less than $75000	61 (31.3%)	83 (31.4%)	407 (12.5%)	551 (14.8%)
$75000 - $99999	40 (20.5%)	43 (16.3%)	480 (14.8%)	563 (15.2%)
$100000 and over	94 (48.2%)	138 (52.3%)	2367 (72.7%)	2599 (70.0%)
Household Income Wave V				
Less than $75000	103 (52.8%)	151 (57.2%)	947 (29.1%)	1201 (32.3%)
$75000 - $99999	45 (23.1%)	43 (16.3%)	635 (19.5%)	723 (19.5%)

```r
long = ready %>% pivot_longer(cols = c(EDUC1,EDUC2,EDUC3),values_to
                              = 'EDUC')


long$HI = NA
for(i in seq(0,11129,3)){
  long$HI[i] = long$w3HI[i]
}

for(i in seq(2,11129,3)){
  long$HI[i] = long$w2HI[i]
}

for(i in seq(1,11129,3)){
  long$HI[i] = long$w1HI[i]
}

long$obs = NA
for(i in seq(0,11129,3)){
  long$obs[i] = long$w3obs[i]
}

for(i in seq(2,11129,3)){
  long$obs[i] = long$w2obs[i]
}

for(i in seq(1,11129,3)){
  long$obs[i] = long$w1obs[i]
}


long$HI = as.factor(
  ifelse(long$HI == 1,"Less than $75000",
         ifelse(
           long$HI == 2, "$75000 - $99999","$100000 and over"
         ))
)

long$age = NA
for(i in seq(0,11129,3)){
  long$age[i] = long$w3age[i]
}

for(i in seq(2,11129,3)){
  long$age[i] = long$w2age[i]
}

for(i in seq(1,11129,3)){
  long$age[i] = long$w1age[i]
}


long = long %>% dplyr::select(-w1obs,-w2obs,-w3obs,-w1HI,-w2HI,-w3HI,-name,-w1age,-w2age,-w3age)
```

```r
long$wave = rep(c(1,2,3),3713)
```

```r
ready %>% ggplot(aes(baselineHI)) +geom_bar(aes(fill = autonomy),position = "fill")
```




```r
long$AID = as.factor(long$AID)

write.csv(long,"tt.csv") # These codes are here because I could not remove a unknown error that was pos
long = read.csv("tt.csv")
```

Analysis

```r
library(geepack)
high = long %>% filter(baselineHI == "$100000 and over")

hightbl1 = geeglm(obs~age+PAEDUC+RACE+BIO_SEX+w2spt+w3spt+autonomy*age+EDUC,corstr = "ar1",id = AID,fam
```

```r
mid = long %>% filter(baselineHI == "$75000 - $99999")
midtbl = geeglm(obs~age+PAEDUC+RACE+BIO_SEX+w2spt+w3spt+autonomy*age+EDUC,corstr = 'ar1',id = AID,data =

low = long %>% filter(baselineHI == "Less than $75000")
lowtbl = geeglm(obs~age+PAEDUC+RACE+BIO_SEX+w2spt+w3spt+autonomy*age+EDUC,corstr="ar1",id = AID,data = l
```

Out puts

```r
tbl_regression(hightbl1)
```

```r
tbl_regression(midtbl)
```

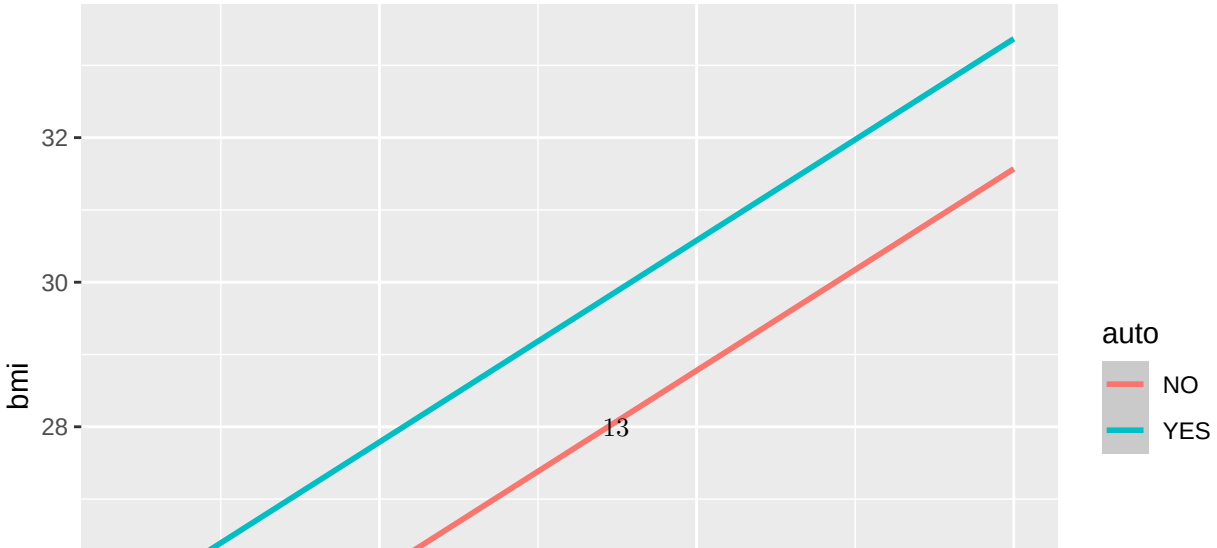
```r
tbl_regression(lowtbl)
```

```r
lowg = data.frame(age = c(12:40), bmi = c(20.40816844 + 0.279 * 12:40 + 1.8,20.40816844 + 0.279 * 12:40)

lowg %>% ggplot(aes(x = age)) + geom_smooth(aes(y = bmi,color = auto))
```

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

Characteristic	Beta	95% CI[1]	p-value
age	0.30	0.28, 0.33	<0.001
PAEDUC			
College Graduate	—	—	
College Graduate +	-0.80	-1.4, -0.18	0.012
High School Graduate	-1.9	-2.7, -1.1	<0.001
Less than High School	-0.49	-1.2, 0.26	0.2
Some College	-0.66	-1.3, -0.03	0.039
RACE			
Asian	—	—	
Black	2.9	1.8, 4.0	<0.001
Hispanic	1.0	-0.07, 2.2	0.066
Indian	5.0	1.7, 8.3	0.003
Multi-Race	1.6	0.27, 3.0	0.019
Other	-0.64	-2.6, 1.3	0.5
White	1.0	0.01, 2.0	0.049
BIO_SEX			
Famale	—	—	
Male	-0.04	-0.41, 0.32	0.8
w2spt	-0.21	-0.32, -0.09	<0.001
w3spt	-0.29	-0.41, -0.17	<0.001
autonomy			
No	—	—	
Yes	0.50	-0.03, 1.0	0.067
EDUC			
College Graduate	—	—	
College Graduate +	0.76	0.34, 1.2	<0.001
High School Graduate	0.22	-0.27, 0.72	0.4
Less than High School	-1.4	-1.9, -0.96	<0.001
Some College	0.09	-0.31, 0.49	0.6
age * autonomy			
age * Yes	-0.02	-0.04, 0.01	0.2

[1]CI = Confidence Interval



Characteristic	Beta	95% CI[1]	p-value
age	0.24	0.16, 0.33	<0.001
PAEDUC			
College Graduate	—	—	
College Graduate +	0.01	-3.1, 3.1	>0.9
High School Graduate	-0.22	-3.3, 2.8	0.9
Less than High School	0.51	-2.6, 3.6	0.7
Some College	-0.40	-3.5, 2.7	0.8
RACE			
Asian	—	—	
Black	4.0	-0.48, 8.4	0.080
Hispanic	2.4	-3.0, 7.9	0.4
Multi-Race	7.0	0.49, 14	0.035
Other	6.3	-0.70, 13	0.078
White	2.2	-2.1, 6.4	0.3
BIO_SEX			
Famale	—	—	
Male	0.66	-0.41, 1.7	0.2
w2spt	-0.11	-0.44, 0.21	0.5
w3spt	-0.27	-0.60, 0.06	0.10
autonomy			
No	—	—	
Yes	-0.20	-2.4, 2.0	0.9
EDUC			
College Graduate	—	—	
College Graduate +	-0.60	-3.1, 1.9	0.6
High School Graduate	-1.5	-4.1, 1.0	0.2
Less than High School	-1.6	-4.3, 1.1	0.2
Some College	-1.5	-3.9, 0.85	0.2
age * autonomy			
age * Yes	0.02	-0.06, 0.11	0.6

[1]CI = Confidence Interval

```
library(sjPlot)
library(sjmisc)
```

```
##
## Attaching package: 'sjmisc'
##
## The following object is masked from 'package:purrr':
##
##     is_empty
##
## The following object is masked from 'package:tidyr':
```

Characteristic	Beta	95% CI[1]	p-value
age	0.28	0.21, 0.34	<0.001
PAEDUC			
College Graduate	—	—	
College Graduate +	-1.9	-4.0, 0.08	0.060
High School Graduate	-2.7	-4.5, -0.85	0.004
Less than High School	-2.7	-4.8, -0.60	0.012
Some College	0.10	-3.0, 3.2	>0.9
RACE			
Asian	—	—	
Black	1.8	-1.4, 5.1	0.3
Hispanic	0.38	-2.3, 3.0	0.8
Multi-Race	3.6	0.55, 6.7	0.021
White	0.82	-0.98, 2.6	0.4
BIO_SEX			
Famale	—	—	
Male	1.2	0.19, 2.1	0.019
w2spt	-0.08	-0.38, 0.22	0.6
w3spt	-0.32	-0.57, -0.06	0.014
autonomy			
No	—	—	
Yes	1.8	0.41, 3.1	0.011
EDUC			
College Graduate	—	—	
College Graduate +	-0.93	-3.7, 1.9	0.5
High School Graduate	-2.1	-4.7, 0.60	0.13
Less than High School	-2.8	-5.7, -0.01	0.050
Some College	-2.7	-5.3, -0.13	0.040
age * autonomy			
age * Yes	-0.06	-0.13, 0.01	0.071

[1]CI = Confidence Interval

```
##
##     replace_na
##
## The following object is masked from 'package:tibble':
##
##     add_case
```

```
library(sjlabelled)
```

```
##
## Attaching package: 'sjlabelled'
##
## The following object is masked from 'package:forcats':
```

```
##
##     as_factor
##
## The following object is masked from 'package:dplyr':
##
##     as_label
##
## The following object is masked from 'package:ggplot2':
##
##     as_label
```

```
allsample = geeglm(obs~age+PAEDUC+RACE+BIO_SEX+w2spt+w3spt+autonomy*age+EDUC,corstr = "ar1",id = AID,fa


summary(allsample)
```

```
##
## Call:
## geeglm(formula = obs ~ age + PAEDUC + RACE + BIO_SEX + w2spt +
##     w3spt + autonomy * age + EDUC, family = gaussian, data = long,
##     id = AID, corstr = "ar1")
##
##  Coefficients:
##                              Estimate  Std.err    Wald Pr(>|W|)
## (Intercept)                 19.06488  0.63498 901.463  < 2e-16 ***
## age                          0.29386  0.01195 604.587  < 2e-16 ***
## PAEDUCCollege Graduate +    -0.86368  0.30645   7.943 0.004828 **
## PAEDUCHigh School Graduate  -2.18916  0.34903  39.340 3.56e-10 ***
## PAEDUCLess than High School -0.83030  0.35040   5.615 0.017807 *
## PAEDUCSome College          -0.65736  0.31000   4.496 0.033966 *
## RACEBlack                    2.99855  0.51005  34.562 4.13e-09 ***
## RACEHispanic                 1.14826  0.53028   4.689 0.030358 *
## RACEIndian                   5.18604  1.67788   9.553 0.001996 **
## RACEMulti-Race               2.05589  0.65068   9.983 0.001580 **
## RACEOther                    0.16698  0.99994   0.028 0.867375
## RACEWhite                    1.02418  0.46811   4.787 0.028676 *
## BIO_SEXMale                  0.07950  0.17110   0.216 0.642173
## w2spt                       -0.19760  0.05359  13.596 0.000227 ***
## w3spt                       -0.29704  0.05531  28.847 7.83e-08 ***
## autonomyYes                  0.55722  0.25242   4.873 0.027279 *
## EDUCCollege Graduate +       0.72596  0.21334  11.579 0.000667 ***
## EDUCHigh School Graduate    -0.07277  0.23182   0.099 0.753609
## EDUCLess than High School   -1.47918  0.23294  40.322 2.15e-10 ***
## EDUCSome College            -0.12215  0.19785   0.381 0.536992
## age:autonomyYes             -0.01800  0.01168   2.377 0.123173
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation structure = ar1
## Estimated Scale Parameters:
##
##             Estimate Std.err
## (Intercept)    38.75   1.062
##   Link = identity
```

```
##
## Estimated Correlation Parameters:
##       Estimate  Std.err
## alpha    0.726 0.007346
## Number of clusters:   3710  Maximum cluster size: 3
```