IPTW Causal Project

```r
library(tidyverse)
```

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

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(readxl)
library(pander)
```

IPTW estimand:

$\Psi(\mathbb{P}_0) = \mathbb{E}_0[\frac{\mathbb{I}(A=1)}{\mathbb{P}_{\not{v}}(A=1|W)}Y]$ - $\mathbb{E}_0[\frac{\mathbb{I}(A=0)}{\mathbb{P}_{\not{v}}(A=0|W)}Y]$

```r
set.seed(252)
ObsData <- read.csv("slpexcov1517.csv")

ObsData <- ObsData %>% dplyr::select(-SEQN, -exminwk, -slphrs, -household, -income, -snoring,
                              -apnea, -bmicat, -smoke, -alcohol, -phq9)

ObsData <- ObsData %>% mutate(A = targetex) %>% mutate(Y = targetslp) %>%
  dplyr::select(-targetex, -targetslp)

ObsData <- na.omit(ObsData)

names(ObsData)
```

```
## [1] "age"       "raceeth"   "educ"      "marital"   "bmi"       "waist"
## [7] "depressed" "A"         "Y"
```

```r
summary(ObsData)
```

```
##       age           raceeth           educ           marital
##  Min.   :20.00   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:31.00   1st Qu.:1.000   1st Qu.:2.000   1st Qu.:1.000
##  Median :43.00   Median :2.000   Median :3.000   Median :2.000
##  Mean   :42.94   Mean   :2.299   Mean   :2.574   Mean   :1.644
```

```
##  3rd Qu.:55.00   3rd Qu.:3.000   3rd Qu.:3.000   3rd Qu.:2.000
##  Max.   :64.00   Max.   :4.000   Max.   :4.000   Max.   :2.000
##       bmi            waist          depressed             A
##  Min.   :15.50   Min.   : 62.30   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:24.82   1st Qu.: 89.83   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :28.35   Median : 99.30   Median :0.00000   Median :0.0000
##  Mean   :29.32   Mean   :101.28   Mean   :0.07017   Mean   :0.4163
##  3rd Qu.:32.80   3rd Qu.:110.60   3rd Qu.:0.00000   3rd Qu.:1.0000
##  Max.   :61.90   Max.   :169.60   Max.   :1.00000   Max.   :1.0000
##        Y
##  Min.   :0.0000
##  1st Qu.:1.0000
##  Median :1.0000
##  Mean   :0.7783
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

1) Create the propensity scores

First fit the logistic regression model:

```
fit<-glm(A~age+factor(raceeth)+factor(educ)+factor(marital)+bmi+waist+factor(depressed),
    family="binomial",data=ObsData)
```

Get propensity scores:

```
prob.1W <- predict(fit, type= "response") #prediced probability of getting the exercise

prob.0W <- 1 - prob.1W #prediced probability of not getting the exercise
```

look at distribution of propensity scores:

$\hat{\mathbb{P}}(A = 1|W_i)$

```
summary(prob.1W) %>% pander
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.05052	0.2721	0.3997	0.4163	0.5506	0.8722

$\hat{\mathbb{P}}(A = 0|W_i)$

```
summary(prob.0W) %>% pander
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.1278	0.4494	0.6003	0.5837	0.7279	0.9495

2) Create the weights:

```
wt1 <- as.numeric(ObsData$A==1)/prob.1W
wt0 <- as.numeric(ObsData$A==0)/prob.0W
```

Look at weights:

$\frac{\mathbb{I}(A_i=1)}{\hat{\mathbb{P}}(A=1|W_i)}$

```
summary(wt1) %>% pander
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	0	0	1.007	1.83	19.79

$\frac{\mathbb{I}(A_i=0)}{\hat{\mathbb{P}}(A=0|W_i)}$

```
summary(wt0) %>% pander
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	0	1.229	0.997	1.586	5.409

3) Estimate:

```
IPTW<- mean( wt1*ObsData$Y) - mean( wt0*ObsData$Y)

IPTW
```

```
## [1] 0.05629606
```

```
unadj<-mean(ObsData$Y[ObsData$A==1])-mean(ObsData$Y[ObsData$A==0])
unadj
```

```
## [1] 0.06442889
```

4) Arbitrarily truncate weights at 10, to see what happens:

First see how many weights are greater than 10:

```
sum(wt1>10)
```

```
## [1] 3
```

```r
sum(wt0>10)
```

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

```r
wt1.trunc<- wt1
wt1.trunc[ wt1.trunc>10] <-10
```

```r
wt0.trunc<- wt0
wt0.trunc[ wt0.trunc>10] <-10
```

IPTW with truncated weights at 10:

```r
mean(wt1.trunc*ObsData$Y) - mean( wt0.trunc*ObsData$Y)
```

```
## [1] 0.05262378
```

What about truncated at 5?

```r
sum(wt1>5)
```

```
## [1] 71
```

```r
sum(wt0>5)
```

```
## [1] 5
```

```r
wt1.trunc5<- wt1
wt1.trunc5[ wt1.trunc5>5] <-5
```

```r
wt0.trunc5<- wt0
wt0.trunc5[ wt0.trunc>5] <-5
```

IPTW with truncated weights at 5:

```r
mean(wt1.trunc5*ObsData$Y) - mean( wt0.trunc5*ObsData$Y)
```

```
## [1] 0.03135125
```

5) Stabilized IPTW estimator - Modified Horwitz Thompson estimator

```r
mean( wt1*ObsData$Y)/mean( wt1) - mean( wt0*ObsData$Y)/mean( wt0)
```

```
## [1] 0.04822426
```