

R Homework Three

Katherine Wolf

Introduction to Causal Inference (PH252D)

April 18, 2020

1 Background and Causal Roadmap.

2 Import and explore the data set RAssign4.csv.

2.1 Import the data set and assign it to object ObsData.

```
library(tidyverse)

ObsData <- read.csv("RAssign4.csv")
```

2.2 Assign the number of riders to n.

```
n <- nrow(ObsData)

n

## [1] 5000
```

2.3 Use the summary function to explore the data.

```
summary(ObsData)

##           W1           W2           A           Y
##  Min.      :1.000   Min.      :0.0000   Min.      :1.000   Min.      :31.49
## 1st Qu.:1.000   1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:36.85
##  Median :2.000   Median :0.0000   Median :4.000   Median :37.83
##  Mean   :2.023   Mean   :0.3486   Mean   :4.085   Mean   :37.92
## 3rd Qu.:3.000   3rd Qu.:1.0000   3rd Qu.:5.000   3rd Qu.:39.32
##  Max.   :3.000   Max.   :1.0000   Max.   :7.000   Max.   :42.45

tail(ObsData)

##           W1 W2 A           Y
## 4995      1  0 4 36.85538
## 4996      3  0 5 40.82337
## 4997      3  0 5 40.82456
## 4998      2  1 3 36.29618
## 4999      3  0 3 38.70285
## 5000      2  1 3 36.28475
```

2.4 Are there certain covariate combinations with limited variability in the exposure (burpees completed)?

```
table(ObsData$W1, ObsData$W2, ObsData$A)
```

```
## , , = 1
##
##
##      0  1
##  1  2  6
##  2  7 20
##  3 25 40
##
## , , = 2
##
##
##      0  1
##  1 40 50
##  2 56 80
##  3 114 132
##
## , , = 3
##
##
##      0  1
##  1 130 124
##  2 174 156
##  3 281 223
##
## , , = 4
##
##
##      0  1
##  1 266 175
##  2 348 172
##  3 324 145
##
## , , = 5
##
##
##      0  1
##  1 318 127
##  2 320 117
##  3 249  63
##
## , , = 6
##
##
##      0  1
##  1 232  57
##  2 156  28
##  3  99  12
##
## , , = 7
##
```

```
##
##      0   1
##    1 69 10
##    2 36  5
##    3 11  1
```

2.4.1 Use the table function to check the number of riders in each exposure-covariate category.

2.4.2 Comment.

do this post-coding

3 IPTW for the statistical estimand equal to the ATE under $\mathcal{M}^{\mathcal{F}^*}$

3.1 We need to estimate the treatment mechanism $\mathbb{P}_0(A|W) = g_0(A|W)$, which is the conditional probability of completing A burpees, given the rider's characteristics.

Implement the following code to estimate the treatment mechanism with multinomial logistic regression. You will need the `nnet` package:

```
library("nnet")

## Warning: package 'nnet' was built under R version 3.6.3

gAW.reg <- multinom(A ~ W1 + W2,
                    data = ObsData)

## # weights: 28 (18 variable)
## initial value 9729.550745
## iter 10 value 8205.693797
## iter 20 value 8032.285228
## final value 8018.904637
## converged
```

3.2 Predict each rider's probability of his/her observed exposure (burpees completed), given his/her covariates $\hat{g}(A_i|W_i)$:

3.2.1 Use the predict function to obtain the predicted probability of each exposure level, given the rider's covariates. Be sure to specify type = "probs".

```
gAW.pred <- predict(gAW.reg, type="probs")
```

3.2.2 Create an empty vector gAW of length n for the predicted probabilities.

```
gAW <- rep(NA, n)
```

3.2.3 Among riders with exposure level $A = 1$, assign the appropriate predicted probability:

```
gAW[ObsData$A==1] <- gAW.pred[ObsData$A==1, "1"]
```

3.2.4 Implement the analogous code for exposure levels $A = 2, \dots, A = 7$:

```
gAW[ObsData$A==2] <- gAW.pred[ObsData$A==2, "2"]
gAW[ObsData$A==3] <- gAW.pred[ObsData$A==3, "3"]
gAW[ObsData$A==4] <- gAW.pred[ObsData$A==4, "4"]
gAW[ObsData$A==5] <- gAW.pred[ObsData$A==5, "5"]
gAW[ObsData$A==6] <- gAW.pred[ObsData$A==6, "6"]
gAW[ObsData$A==7] <- gAW.pred[ObsData$A==7, "7"]
```

3.2.5 Use the `summary` function to examine the distribution of predicted probabilities. Any cause for concern?

```
summary(gAW)
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.002224 0.175956 0.251250 0.228126 0.299407 0.341685
```

talk about cause for concern after programming

3.3 Create the vector `wt` as the inverse of the predicted probabilities. Use the `summary` function to examine the distribution of weights. Comment on the distribution of weights.

```
wt <- 1/gAW
summary(wt)
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  2.927   3.340   3.980   6.934   5.683  449.690
```

comment after programming

3.4 Evaluate the IPTW estimand:

$$\hat{\Psi}_{IPTW}(\mathbb{P}_n) = \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{I}(A_i = 7)}{\hat{g}(A_i|W_i)} Y_i - \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{I}(A_i = 1)}{\hat{g}(A_i|W_i)} Y_i$$

The first quantity is the weighted mean outcome, where riders completing $A_i = 7$ burpees receive weight $\frac{1}{\hat{g}(A_i=7|W_i)}$ and riders completing $A_i \neq 7$ burpees receive weight 0. The second quantity is the weighted mean outcome, where riders completing $A_i = 1$ burpees receive weight $\frac{1}{\hat{g}(A_i=1|W_i)}$ and riders completing $A_i \neq 1$ burpees receive weight 0.

```
iptw.estimand <- mean(wt*as.numeric(ObsData$A==7)*ObsData$Y) -
  mean(wt*as.numeric(ObsData$A==1)*ObsData$Y)

iptw.estimand
## [1] 7.531834
```

3.5 Implement the stabilized IPTW estimator (a.k.a. the modified Horvitz-Thompson estimator):

$$\hat{\Psi}_{st.IPTW} = \frac{\frac{1}{n} \sum_{i=1}^n \frac{\mathbb{I}(A_i=7)}{\hat{g}(A_i|W_i)} Y_i}{\frac{1}{n} \sum_{i=1}^n \frac{\mathbb{I}(A_i=7)}{\hat{g}(A_i|W_i)}} - \frac{\frac{1}{n} \sum_{i=1}^n \frac{\mathbb{I}(A_i=1)}{\hat{g}(A_i|W_i)} Y_i}{\frac{1}{n} \sum_{i=1}^n \frac{\mathbb{I}(A_i=1)}{\hat{g}(A_i|W_i)}}$$

```
iptw.ht.estimand <-  
  mean(wt*as.numeric(ObsData$A==7)*ObsData$Y)/  
  mean(wt*as.numeric(ObsData$A==7)) -  
  mean(wt*as.numeric(ObsData$A==1)*ObsData$Y)/  
  mean(wt*as.numeric(ObsData$A==1))
```

3.6 Interpret the point estimates.

do this after programming

4 IPTW and Marginal Structural Models

4.1 IPTW for the MSM parameter without stabilized weights

- 4.1.1 Estimate the treatment mechanism $\mathbb{P}_0(A|W) = g_0(A|W)$, which is the conditional probability of completing A burpees, given the rider's characteristics. Use multinomial logistic regression. *Hint: We already did this! Skip to the next step.*
- 4.1.2 Predict each rider's probability of her observed exposure (burpees completed), given his/her covariates $\hat{g}(A_i|W_i)$. *Hint: We already did this! Skip to the next step.*
- 4.1.3 Create the vector `wt` as the inverse of the predicted probabilities. *Hint: We already did this! Skip to the next step.*
- 4.1.4 Estimate the parameters corresponding to the MSM by regressing the observed outcome Y on the exposure A according to $m(a|\beta)$. You must specify the weights and the data.

```
iptw.msm <- glm(Y ~ A ,  
               weights = wt,  
               data = ObsData)  
  
iptw.msm$coef  
  
## (Intercept)          A  
## 33.084091    1.095612  
  
summary(iptw.msm)  
  
##  
## Call:  
## glm(formula = Y ~ A, data = ObsData, weights = wt)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max   
## -49.508  -1.338   1.428   2.470  15.694   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) 33.08409    0.04657   710.4   <2e-16 ***  
## A           1.09561    0.01043   105.1   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 14.87555)
##
##      Null deviance: 238519  on 4999  degrees of freedom
## Residual deviance:  74348  on 4998  degrees of freedom
## AIC: 19673
##
## Number of Fisher Scoring iterations: 2
```

4.1.5 Interpret the results.

do after programming QUESTION: INCLUDE COVARIATES?

4.2 IPTW for a MSM parameter with stabilized weights

4.3 Implement IPTW for a MSM parameter with stabilized weights

4.3.1 Estimate the treatment mechanism $g_0(A|W) = \mathbb{P}_0(A|W)$. *Hint: We already did this! Skip to the next step.*

4.3.2 Predict the probability of the observed exposure for each rider gAW . *Hint: We already did this! Skip to the next step.*

4.3.3 Create the stabilized weights $wt.MSM$:

$$\frac{\hat{g}^*(A)}{\hat{g}(A|W)}, \text{ where } \hat{g}^*(A) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(A_i = a)$$

1. Create empty vector gA of length n for the numerator of the weights.

```
gA <- rep(NA, n)
```

2. Index the vector gA by exposure status and assign the appropriate estimated probability. *Hint: For riders completing $A = 1$ burpee, the numerator $g^*(A)$ is the observed proportion with $A = 1$.*

```
gA[ObsData$A==1] <- mean(ObsData$A==1)
gA[ObsData$A==2] <- mean(ObsData$A==2)
gA[ObsData$A==3] <- mean(ObsData$A==3)
gA[ObsData$A==4] <- mean(ObsData$A==4)
gA[ObsData$A==5] <- mean(ObsData$A==5)
gA[ObsData$A==6] <- mean(ObsData$A==6)
gA[ObsData$A==7] <- mean(ObsData$A==7)
```

3. Create the stabilized weights:

```
wt.MSM <- gA/gAW
```

4. Comment on the distribution of the stabilized weights.

[do this post-programming](#)

4.3.4 Estimate the parameters corresponding to the MSM by regressing the observed outcome Y on the exposure A . You must specify the weights and the data.

```
iptw.msm.st <- glm(Y ~ A ,
                  weights = wt.MSM,
                  data = ObsData)

iptw.msm.st$coef

## (Intercept)          A
##   33.671481    1.025057

summary(iptw.msm.st)

##
## Call:
## glm(formula = Y ~ A, data = ObsData, weights = wt.MSM)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -8.5159  -0.8660   0.3280   0.8466   2.4091
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.67148    0.06476  519.97  <2e-16 ***
## A            1.02506    0.01513   67.75  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.945574)
##
##      Null deviance: 18654  on 4999  degrees of freedom
## Residual deviance:  9724  on 4998  degrees of freedom
## AIC: 17901
##
## Number of Fisher Scoring iterations: 2
```

4.3.5 Are the estimated coefficients the same? Briefly discuss.

[do this post-programming](#)

5 Improving the IPTW estimator and the G-computation formula

5.1 Run the code given in `Rassign4_modifiedIPTW.R` and report how the standard IPTW and modified Horvitz-Thompson estimators perform in terms of bias, variance, and MSE. Which estimator would you use in practice?

```
source("Rassign4_ipwtModified.R")

## [1] "The estimators have (estimated) bias:"
##      IPTW      Modified HT    simple.mean      my.est      my.est.2
## 4.990738e-01 2.180333e-04 -2.279129e-02 -7.738690e-06 -4.155064e-01
## [1] "The estimators have (estimated) variance:"
##      IPTW      Modified HT    simple.mean      my.est      my.est.2
## 2.179920e+03 7.772966e-04 2.408178e-04 7.702404e-04 8.702269e-05
## [1] "The estimators have (estimated) MSE:"
##      IPTW      Modified HT    simple.mean      my.est      my.est.2
## 2.180170e+03 7.773441e-04 7.602606e-04 7.702405e-04 1.727326e-01
```

The modified Horvitz-Thompson IPTW estimator has lower bias, variance, and mean squared error for this sample than the standard IPTW estimator, and I would likely use it in practice. That said, with a large enough sample, both estimators should be unbiased, but I would still likely use the Horvitz-Thompson estimator because it weights uncommon exposure-covariate combinations in the data (i.e., covariate combinations with small numbers of either exposed or unexposed observations) less heavily than standard IPTW estimator, which makes it less sensitive to theoretical and practical positivity violations, generally reduces its variance, and leads to better final sample performance than that of the standard IPTW estimator.

5.2 Look at the IPTW column in the `est` matrix. What do you notice about the IPTW estimates across these 2000 Monte Carlo draws?

```
summary(est[, "IPTW"])

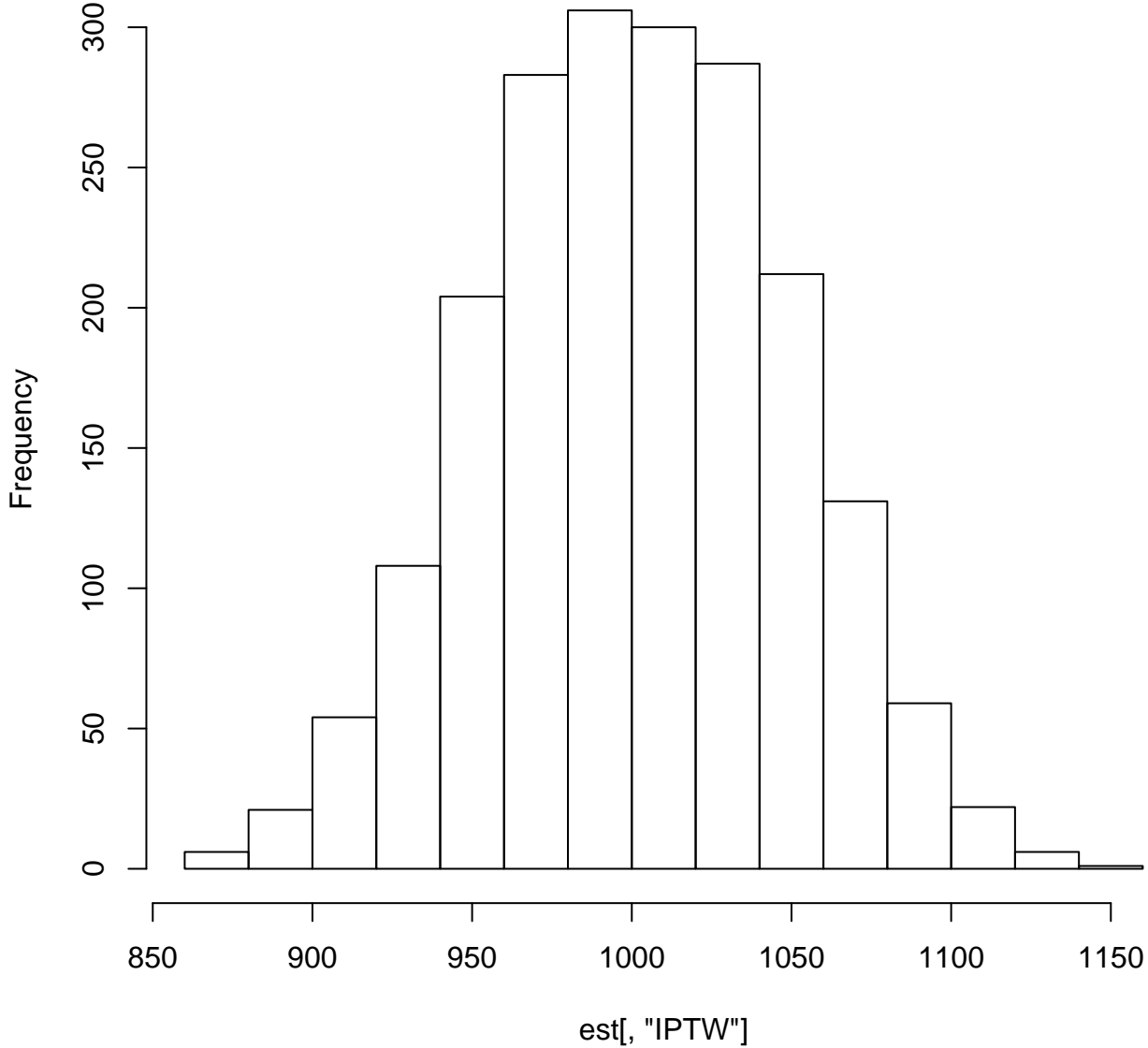
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   863.0   968.1  1000.6  1001.1  1034.4  1145.7

var(est[, "IPTW"])

## [1] 2179.92

hist(est[, "IPTW"])
```


Histogram of est[, "IPTW"]



Y can only take on one of two values, $Y = 1000$ or $Y = 1001$, but the IPTW estimates include Y values ranging from 863.02 to 1145.69.

5.3 What is the variance of a random variable X with $Pr(X = 0) = 1/2$ and $Pr(X = 1) = 1/2$?

Such a random variable is a Bernoulli random variable with $p = Pr(X = 1) = 1/2$ and $q = Pr(X = 0) = 1/2$. Thus the mean of X is $E[X] = \frac{1}{2} * 1 + \frac{1}{2} * 0 = \frac{1}{2}$, and the expected value of X^2 is $E[X^2] = \sum_{x \in \text{range}(X)} x^2(\mathbb{P}_x) = 1^2 * \frac{1}{2} + 0^2 * \frac{1}{2} = \frac{1}{2}$. Since variance of a random variable is by definition $Var[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$, then the variance of our Bernoulli random variable friend X is $Var[X] = E[X^2] - (E[X])^2 = \frac{1}{2} - (\frac{1}{2})^2 = \frac{1}{4}$.

5.4 What is the variance of a random variable X_2 with $Pr(X_2 = 0) = 1/2$ and $Pr(X_2 = 1000) = 1/2$?

The mean of such a variable would be $E[X_2] = \sum_{x_2 \in \text{range}(X_2)} x_2(\mathbb{P}_{x_2}) = \frac{1}{2} * 0 + \frac{1}{2} * 1000 = 500$.

The expected value of the square of such a variable would be $E[X^2] = \sum_{x \in \text{range}(X)} x^2 (\mathbb{P}_x) = 0^2 * \frac{1}{2} + 1000^2 * \frac{1}{2} = 5 \times 10^5$. Thus the variance of such a variable would be $\text{Var}[X] = E[X^2] - (E[X])^2 = 5 \times 10^5 - 500^2 = 2.5 \times 10^5$.

5.5 Can you think of how the above two calculations could be relevant to improving the IPTW estimator in this problem?

If we find a way to reduce the range of the predicted Y values that the estimator averages to estimate $\psi(\mathbb{P}_0)$, the resulting estimator will be more efficient, i.e., its estimate will have a smaller variance, which could possibly reduce the ultimate mean squared error.

5.6 *Bonus:* Write down an estimator $\hat{\Psi}_{my.est}$ which applies the ideas of the previous three questions into an estimator. There's no need to give the best possible estimator, but you should give an estimator that outperforms the IPTW estimator by a significant margin (i.e. does as or almost as well as the modified Horvitz-Thompson estimator in terms of bias/variance/MSE).

Below is the IPTW estimator stabilized by subtracting the mean of the observed Y values from each Y_i and then adding that mean back at the end before calculating the overall mean.

$$\hat{\psi}_{katie}(\mathbb{P}_n) = \frac{1}{n} \sum_{i=1}^n \left[\frac{A_i(Y_i - \sum_{i=1}^n (Y_i))}{g_0(1|W_i)} + \sum_{i=1}^n (Y_i) \right]$$

5.7 *Bonus:* Code your estimator and replace the NA on the line `my.est = NA` with the estimator you defined in the previous question. Report the bias/variance/MSE of your estimator over the 2000 Monte Carlo draws.

```
source("Rassign4_iprwModified.R")

## [1] "The estimators have (estimated) bias:"
##      IPTW      Modified HT      simple.mean      my.est      my.est.2
## 4.990738e-01 2.180333e-04 -2.279129e-02 -7.738690e-06 -4.155064e-01
## [1] "The estimators have (estimated) variance:"
##      IPTW      Modified HT      simple.mean      my.est      my.est.2
## 2.179920e+03 7.772966e-04 2.408178e-04 7.702404e-04 8.702269e-05
## [1] "The estimators have (estimated) MSE:"
##      IPTW      Modified HT      simple.mean      my.est      my.est.2
## 2.180170e+03 7.773441e-04 7.602606e-04 7.702405e-04 1.727326e-01
```

My estimator beats both the standard IPTW and modified modified Horvitz-Thompson estimator in bias, variance, and mean squared error! It fails, however, to beat taking a simple mean of the observed Y values in terms of mean square error.